AD_____
(Leave blank)

Award Number: W81XWH-07-1-0404

TITLE:
Mechanisms underlying the breast cancer susceptibility locus *Mcs5a*

PRINCIPAL INVESTIGATOR: Bart M. G. Smits, Ph.D.

CONTRACTING ORGANIZATION:
University of Wisconsin
Madison, WI 53706

REPORT DATE: July 2010

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: (Check one)

V Approved for public release; distribution unlimited

Distribution limited to U.S. Government agencies only;
report contains proprietary information

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-07-2010	Annual Summary	1 July 2007 – 30 June 2010

4. TITLE AND SUBTITLE

Mechanisms underlying the breast cancer susceptibility locus *Mcs5a*

5a. CONTRACT NUMBER
W81XWH-07-1-0404

5b. GRANT NUMBER
BC061963

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Bart M. G. Smits
Email: bsmits@wisc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Wisconsin
Madison, WI 53706

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Material Command
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

For low-penetrance breast cancer risk alleles it is currently unknown how they lead to predisposition. Here, we study the *Mcs5a* locus that is associated with breast cancer risk in rats and humans. In our rat model we show that the presence of the resistant genotype of two components of the locus (*Mcs5a1, Mcs5a2*) down regulates the expression of the *Fbxo10* gene in the T cells and that this reduced expression is associated with reduced mammary tumor multiplicity. We show that genetic elements in *Mcs5a1* and *Mcs5a2* are physically close to each other in the nuclear space. The spatial organization of the locus in primary T cells is conserved between rat and human. We demonstrate that the function of the non-coding *Mcs5a* locus likely is repressive gene regulation. We present a model that begins to explain how the *Fbxo10* gene could be regulated in T cells.

15. SUBJECT TERMS

mammary carcinogenesis, susceptibility, rat, human, breast cancer risk, *Fbxo10*, expression, T cells, chromatin organization, 3C, genetic elements

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	32	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Page

INTRODUCTION

Breast cancer susceptibility is a complex, polygenic trait, in which the cumulative effects of low penetrance, high population frequency, risk-altering genetic variants (modifiers) determine the heritable fraction. To be able to construct genetic risk profiles and population-based intervention programs directed to those at highest risk, it is important to identify as many risk alleles as possible. Using whole-genome linkage studies in inbred rat models that vary in susceptibility to carcinogen (DMBA; 7,12-dimethylbenz(a)anthracene)-induced mammary cancer, we found mammary carcinogenesis susceptibility QTL *Mcs5* (Samuelson et al., 2003). Using congenic recombinant inbred lines that have small pieces of the resistant genome introgressed in the susceptible background, *Mcs5* was found to contain at least three distinct loci (*Mcs5a*-*c*) (Samuelson et al., 2005). *Mcs5a* has been mapped to ultra-fine resolution and was found to be a compound QTL, consisting of two loci (*Mcs5a1*, ~ 30 Kb; *Mcs5a2*, ~80 Kb) that synthetically interact only in *cis* (on the same chromosome) to confer resistance (Samuelson et al., 2007). Human *MCS5A* has essentially the same genetic features as rat *Mcs5a*. Interestingly, in two population-based case-control studies (~12,000 women), the minor alleles of a SNP (single nucleotide polymorphism) in human *MCS5A1* and a SNP in human *MCS5A2* associate significantly with an altered breast cancer risk (Samuelson et al., 2007). These SNPs could either be causative themselves, or be a marker for the causative SNP. This human association study clearly demonstrates the utility of rat models to identify unbiased potential human breast cancer candidates.

Since *Mcs5a* is entirely non-coding, the causative genetic elements will likely involve transcriptional regulation. All genes within 0.5 Mb flanking the QTL are expressed at similar levels in the mammary glands in susceptible and resistant congenic animals (Samuelson et al., 2007). However, *Fbxo10* and *Frmpd1*, the genes transcriptionally starting off in *Mcs5a*, are differentially expressed in thymus and spleen, respectively. However, only the expression level of *Fbxo10* in the thymus is correlated with mammary carcinogenesis susceptibility (Smits et al., submitted). *Mcs5a1* and *Mcs5a2* also need to be both present to reduce the expression in the thymus. Flow cytometry experiments revealed that the *Fbxo10* differential expression is limited to T cells (Smits et al., submitted). In addition, a mammary gland transplantation assay indicated that there is a host effect on the mammary carcinogenesis phenotype mediated by *Mcs5a*, suggesting a mammary cell-non autonomous effect of the *Mcs5a* locus on mammary carcinogenesis susceptibility (Smits et al., submitted).

We hypothesize that in T cells, genetic elements in *Mcs5a1* and *Mcs5a2* are looping over to physically interact in order to regulate the expression of the *Fbxo10* gene.

BODY

Training

Lab Meetings and Seminars at the University of Wisconsin (SoW Task 1)

As part of my postdoctoral training, I participated by attending and presenting in the Gould lab meetings and in the student/postdoc seminar series organized by the McArdle Lab for Cancer Research. On a weekly basis, I attended seminars given by invited specialists on diverse cancer biology related topics, including transcriptional regulation, biostatistics, genetics, genomics, and more.

Visit Dr. Job Dekker's lab (SoW Task 2)

The chromatin conformation capture (3C) technology is a crucial procedure for understanding how the *Mcs5a* locus is organized in the nucleus to regulate the expression of *Fbxo10*. The 3C assay was invented by Dr. Job Dekker in 2002 (Dekker et al., 2002). I visited Dr. Dekker's lab at the University of Massachusetts Medical School (Worcester, MA) to obtain hands-on training in the 3C assay. Following the visit, I have successfully implemented the 3C assay in the Gould lab. Using 3C, I profiled the *Mcs5a* region in various rat cell types and human cell lines.

Scientific Meetings (SoW Task 3)

I took part in three international scientific meetings, namely Keystone Symposia 'Complex Traits: Biological and Therapeutic Insights', Santa Fe, NM, held February 29 – March 5, 2008, 'Chromatin Dynamics and Higher-Order organization', Coeur D'Alene, ID, held February 25 – March 2, 2009, and Cold Spring Harbor Laboratories Meeting 'Rat Genomics and Models', Cold Spring Harbor, NY, held December 2 - December 5, 2010.

I participated with a poster presentation in the Era of Hope DoD BCRPM meeting, Baltimore, MD, held June 25 – 28, 2008. I participated in two BCRP LINKS meetings, in 2009 and 2010.

Mentoring Committee (SoW Task 4)

Although a formal meeting with the entire mentoring committee has not taken place, I had regular discussions with the members separately. I had regular discussions with my primary mentor, Dr. Michael Gould, at least once a week. In 2008, and 2009, I presented my work at the Transcriptional Mechanisms seminar series organized by Dr. Emery Bresnick, which was followed by a discussion. Discussions with Dr. William Dove and Dr. Sündüz Keles took place regularly, when needed.

Research

The 3C assay (SoW Task 1)

To identify a physical interaction between genetic elements in *Mcs5a1* and *Mcs5a2*, implementation of the chromatin conformation capture (3C) assay is essential. In collaboration with the lab of Dr. Job Dekker, who invented the 3C assay in 2002 and has a broad experience in using it (Dekker, 2006), the 3C assay was established in the Gould lab (SoW Task 1a). To capture chromosomal interactions, cells are fixed using formaldehyde. The extracted fixed chromatin is digested with a restriction enzyme and religated in a strongly dilute fashion. In this procedure the ligation of genetic elements that were glued together by formaldehyde fixation is favored over ligation of random elements. Following reversal of the crosslinks, the ligation frequency of two elements of interest is determined quantitatively. The measurements will be relative to a fully digested and randomly ligated control template containing all restriction fragments of interest in equal molarity. To investigate the *Mcs5a1-Mcs5a2* interaction a fixed fragment in *Mcs5a1* was chosen and the relative interaction frequency to all restriction fragments in *Mcs5a2* was determined. This results in a regional profile in which fragments close to the fixed fragments give a high relative interaction frequency, due to random ligation events. Such random events decrease with increasing genomic distance. Local peaks in the profile are indicative of a physical interaction.

The 3C assay was applied to our rat models to address three fundamental questions about the structural organization of the *Mcs5a* locus: 1. Does the structural organization support a physical interaction between an element in *Mcs5a1* and an element in *Mcs5a2*? 2. Does the susceptible or resistant genotype have an effect on the structural organization of *Mcs5a*? 3. Is the structural organization of *Mcs5a* different between various tissues / cell types?

To answer the first question, if *Mcs5a1* is actually in close proximity to *Mcs5a2*, the 3C assay was used on (splenic) T-lymphocytes of the susceptible WF.WKy strain. This cell type has been shown to differentially express the *Fbxo10* gene and is considered the cell type of action, as described below (SoW Task 2b). The 3C assay was performed with three different fixed fragments in *Mcs5a1* that were probed for interactions with all working restrictions fragments in *Mcs5a2*. The fixed fragments in *Mcs5a1* were chosen to be close to the *Fbxo10* promoter and putative regulatory elements. Figures 1a and 1b show the chromatin profiles for the first two fixed fragments, closest to the *Fbxo10* putative proximal promoter, as determined by start site analysis, described later. These two fixed fragments do not show outstanding interactions with any elements in *Mcs5a2*. In figure 1c the fixed fragment is slightly shifted towards putative upstream regulatory elements of the *Fbxo10* gene. Probing all *Mcs5a2* restriction fragments with this fixed fragment yielded locally enhanced interaction frequencies with at least two elements in *Mcs5a2*, close to the *Mcs5a1/Mcs5a2* border. This 3-way interaction was confirmed by using one of the two *Mcs5a2* interacting elements as the fixed fragment and scanning both ways (Fig. 1d).

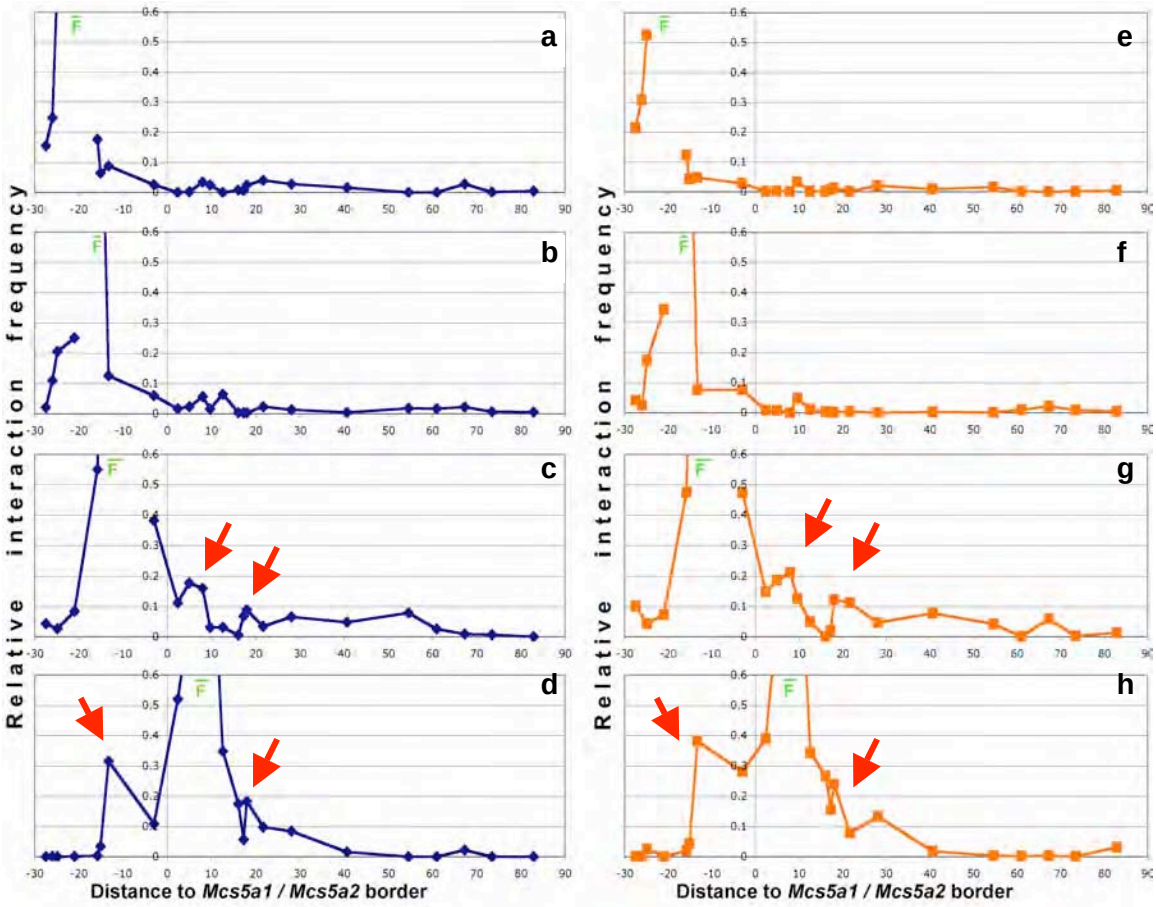

Figure 1: 3C profiles of the rat *Mcs5a* locus in splenic T cells of susceptible line WF.WKy (a-d) and resistant congenic animals WW (e-h). The fixed fragments used are indicated by a green bar. Each point is the average of at least three measurements on 3C template pools of six rats per genotype. Red arrows indicate areas of potential looping.

To investigate the second question concerning the effect of the susceptible or resistant genotype on putative looping, rats of the susceptible control line (WF.WKy), susceptible congenic lines (*Mcs5a1* and *Mcs5a2*), and the resistant congenic line (WW) were used (SoW Task 1b). Figures 1e-h show the chromatin profiles of the *Mcs5a* locus of splenic T-lymphocytes of the resistant congenic line WW, as determined by 3C using the same fixed fragments as in figures 1a-d. The chromatin profiles of T-lymphocytes of the susceptible line WF.WKy and the resistant line WW did not differ significantly. Accordingly, the profiles for the susceptible congenic lines *Mcs5a1* and *Mcs5a2* were not found to differ either (not shown), leading to the conclusion that the structural organization of the *Mcs5a* locus is not affected by the susceptible or resistant genotype.

By determining the structural organization of *Mcs5a* in different cell types in our rat models, the third fundamental question was answered. Figures 2a-c display the 3C profiles for splenic non-T cells, splenic T cells, and mammary gland, respectively. The splenic non-T cell population primarily contains B cells, and monocytes. Both these cell types, and the mammary gland have been shown not to differentially express the *Fbxo10* gene, which is in contrast to T-lymphocytes, as described later (SoW Task 2b). Regardless of the *Fbxo10* expression differences, the genetic elements in the 3-way interaction are identical in these three cell types, although the signal intensity is enhanced in the mammary gland. This result led to the conclusion that the chromatin

structure of *Mcs5a* may not be a direct determinant of the differential expression of the *Fbxo10* gene seen in T-lymphocytes, but the structure may facilitate transcriptional regulation by certain genetic elements yet to be determined.

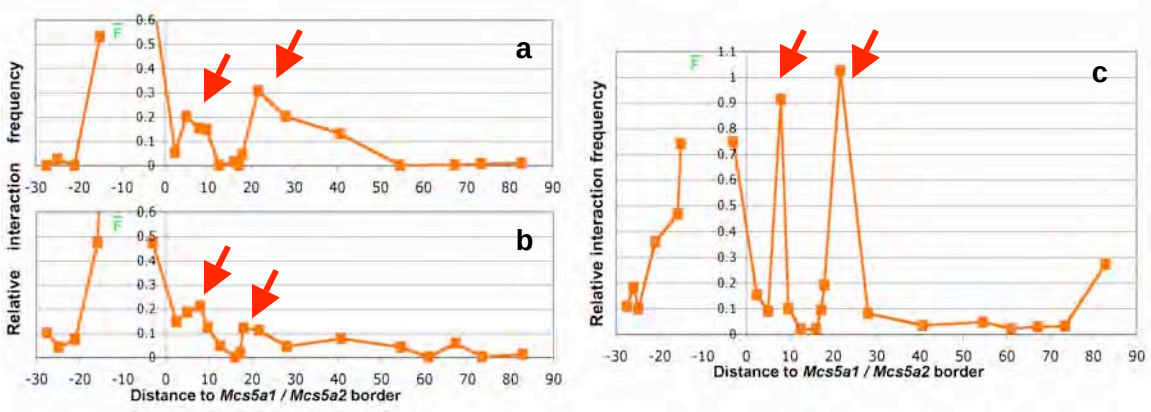

Figure 2: 3C profiles of the rat *Mcs5a* of resistant congenic animals WW in splenic non-T lymphocytes (a), splenic T-lymphocytes (b), and mammary gland (c). The fixed fragments used are indicated by a green bar. Again each point is the average of at least three measurements on 3C template pools of 4-6 rats per genotype. Red arrows indicate areas of potential looping.

Finally, the 3C technology was applied to various human cell lines (SoW Task 1c). Human *MCS5A* has the same genetic features as rat *Mcs5a*. In both human orthologous regions of *Mcs5a1* and *Mcs5a2*, novel breast cancer risk alleles were identified in a large case-control study (Samuelson et al., 2007). The question now becomes: Besides the sequence, the genetic features, and the association with breast cancer risk, is the structural organization also conserved between rat and human *Mcs5a*? Figures 3a-c display the chromatin structure of *MCS5A* of a cervical carcinoma cell line (HeLa), a mammary carcinoma cell line (MCF-7), and a leukemic T-lymphocyte cell line (JURKAT). These profiles were determined using a fixed fragment that includes the promoter of the *FBXO10* gene. In the rat, no clear interactions were detected using fixed fragments with the *Fbxo10* promoter, as described above (Fig. 1a,b,e,f).

It turns out that in all three cell lines many interactions were picked up, indicative of more condensed chromatin, possibly due to the cancerous nature of the cell lines (Holloway and Oakford, 2007). When primary T-lymphocytes were used, the human *MCS5A* chromatin profile did not show any obvious interactions (Fig. 3d), which resembles the profile in rat T-lymphocytes (Fig. 1a,b,e,f). Similarly, when the fixed fragment was shifted towards putative upstream regulatory elements, two areas of enhanced local interaction frequency close to the *MCS5A1/MCS5A2* border could be picked up again, just like in rat T-lymphocytes. Primary T-lymphocytes nicely reflected the structural organization of *MCS5A*, however, human cell lines of cancerous origin do not fully reflect the structural organization of the *MCS5A* locus. Therefore, these cell lines might not be suitable to model the gene regulatory properties of the *MCS5A* locus.

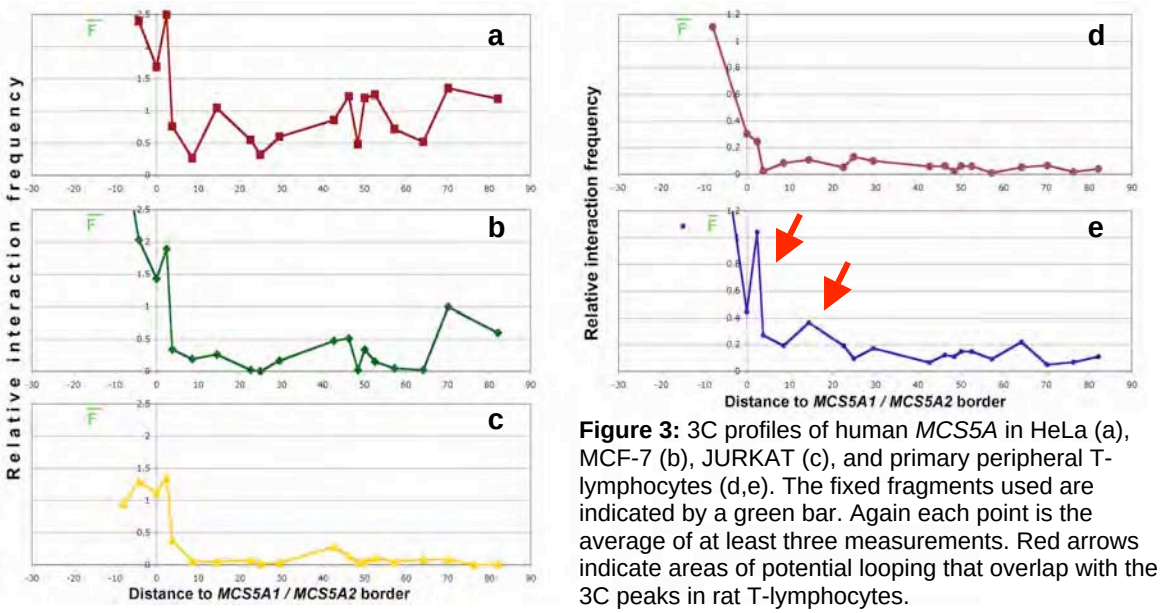

Figure 3: 3C profiles of human *MCS5A* in HeLa (a), MCF-7 (b), JURKAT (c), and primary peripheral T-lymphocytes (d,e). The fixed fragments used are indicated by a green bar. Again each point is the average of at least three measurements. Red arrows indicate areas of potential looping that overlap with the 3C peaks in rat T-lymphocytes.

The CCCTC-binding factor (CTCF) protein is widely known for its role as a vertebrate insulator protein (Bell et al., 1999). Additionally, CTCF has been shown to be essential for long-distance enhancer-promoter looping (Splinter et al., 2006). Recently, another protein, cohesin, has also been shown to be essential for long-range chromatin looping in the developmentally controlled *IFNG* locus (Hadjur et al., 2009). We sought to investigate whether CTCF/cohesin binding could also underlie the observed higher-order chromatin structure of *MCS5A/Mcs5a*. I examined binding of CTCF/cohesin to the fragments involved in the 3-way chromatin interaction across the *MCS5A* locus in both human and rat T-lymphocytes using the chromatin immunoprecipitation (ChIP) assay. Briefly, T-lymphocytes (human JURKAT, or rat splenic T-lymphocytes) were formaldehyde fixed and sonicated to shear the chromatin. The sheared chromatin extract was incubated with a monoclonal antibody against CTCF, RAD21 (Cohesin), or IgG. After collecting CTCF antibody-bound chromatin complexes, the DNA was recovered using phenol-chloroform extractions. Enrichment of certain DNA fragments in the CTCF or RAD21 (Cohesin) antibody-collected sample versus a mock antibody (IgG)-collected sample was determined using a quantitative PCR method.

Using ChIP, binding of both CTCF and cohesin to all three interacting fragments was confirmed (Fig. 4a). Binding of CTCF and cohesin to the orthologous sites in the rat *Mcs5a* locus was also identified (Fig. 4a). No evidence of CTCF or cohesin binding was found in locations outside of the looping fragments (Fig. 4a), suggesting that CTCF and cohesin play a role in the higher-order chromatin structure of human and rat *MCS5A/Mcs5a*.

MCS5A harbors a CTCF/cohesin-driven insulator-like chromatin loop, which spatially isolates the *TOMM5* gene and its flanking regulatory region by locating them within the looped chromatin structure (Fig. 4b). The risk alleles in *MCS5A1* and *MCS5A2* are located on either sides of the loop, in closer physical proximity than may be implied from a linear genome view.

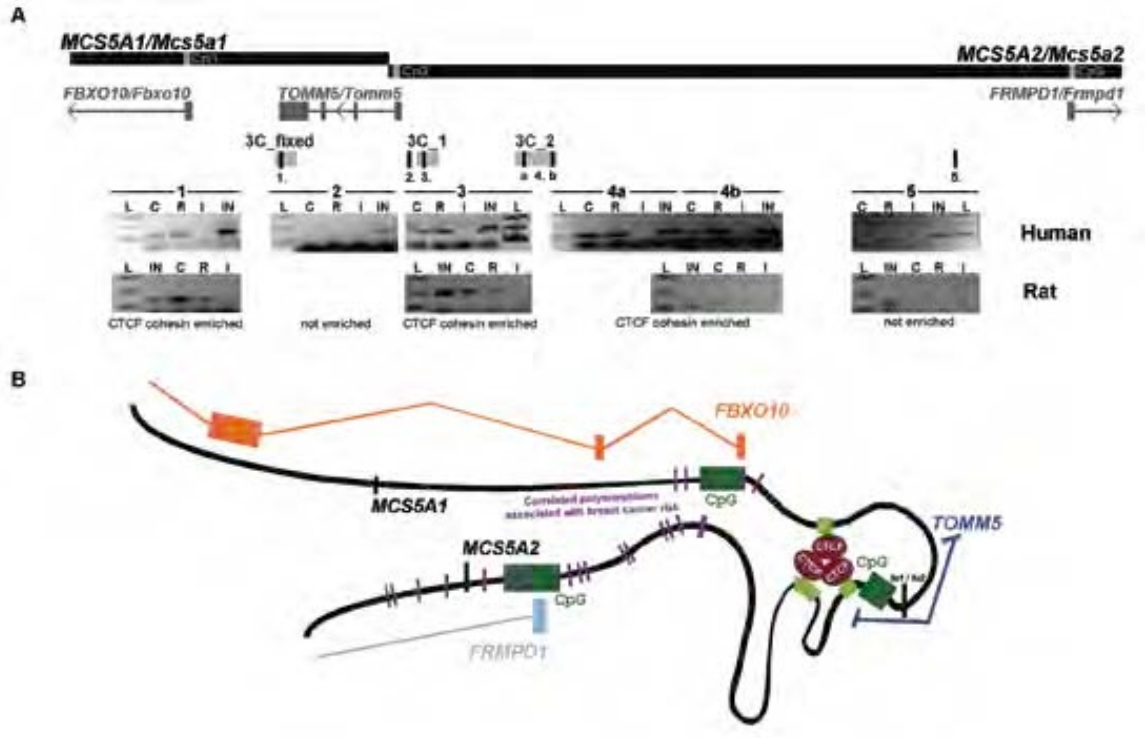

Figure 4: a) CTCF and cohesin chromatin immunoprecipitation (ChIP) assay in the *MCS5A/Mcs5a* locus. The *MCS5A1/Mcs5a1* and *MCS5A2/Mcs5a2* loci are depicted as black lines. The light grey bars within *MCS5A1* and *MCS5A2* are the CpG islands associated with the promoters of the *FBXO10/Fbxo10*, *TOMM5/Tomm5*, and *FRMPD1/Frmpd1* genes respectively. The locations of the three interacting elements in 3C are indicated by light grey blocks. Several locations within and outside the interacting fragments were analyzed by PCR on CTCF (C), cohesin (R; Rad21), and IgG (I; negative control) antibody immunoprecipitated chromatin samples, and an input (IN, positive control) sample, prepared from JURKAT cells (human) or primary rat splenic T-lymphocytes (rat). Each gel image is accompanied by a 100-bp DNA ladder (L) of which the lower three bands (100, 200, 300 bp) are shown. **b)** A model of the *MCS5A* locus in a folded configuration. The *FBXO10* transcript is displayed in orange. The *TOMM5* transcript is indicated in dark blue. The *FRMPD1* transcript is shown in light blue. The CpG islands associated with their promoters are indicated in dark green. The correlated polymorphisms that associate with breast cancer risk are depicted as purple bars. The interacting fragments in the 3C assay are shown in light green.

Regulation of gene expression (SoW Task 2)

The next step is to understand how the locus regulates gene expression that ultimately predisposes to breast cancer. Previous expression analysis of all genes within 1 Mb of the *Mcs5a* locus in the rat mammary gland of susceptible (WF.WKy) and resistant congenic (WW) animals yielded no expression differences (Samuelson et al., 2007). A co-worker in the Gould lab, Dr. David Samuelson, proceeded with expression analysis in other tissues of WF.WKy and WW animals and found the *Fbxo10* gene to be differentially expressed in the thymus and the *Frmpd1* gene in the spleen (SoW Task 2b). When these two genes were profiled in the thymus and spleen of the two susceptible congenic lines (*Mcs5a1, Mcs5a2*) that just have *Mcs5a1* or *Mcs5a*2 of the resistant genotype, only the expression level of *Fbxo10* in the thymus appeared to be correlated with the mammary carcinogenesis susceptibility phenotype (Fig. 5a). In other

words, the *Mcs5a1-Mcs5a2* interaction is required for both down regulation of thymic *Fbxo10* expression and reduced tumor multiplicity in our carcinogenesis model.
The thymus consists mainly of T-lymphocytes that could be expressing the CD4 receptor (CD4+), the CD8 receptor (CD8+), both receptors (double positive), or none of the receptors (double negative). Following flow cytometric separation of these cell types, *Fbxo10* was found to be differentially expressed in single positive CD4+, single positive CD8+, and double positive thymocytes (Fig. 5b). When isolated from the spleen, CD3+ T lymphocytes persisted in their differential *Fbxo10* expression, whereas other cell types isolated from the spleen did not (Fig. 5c). At this point the T lymphocytes are considered the cell type of action (SoW Task 2b).

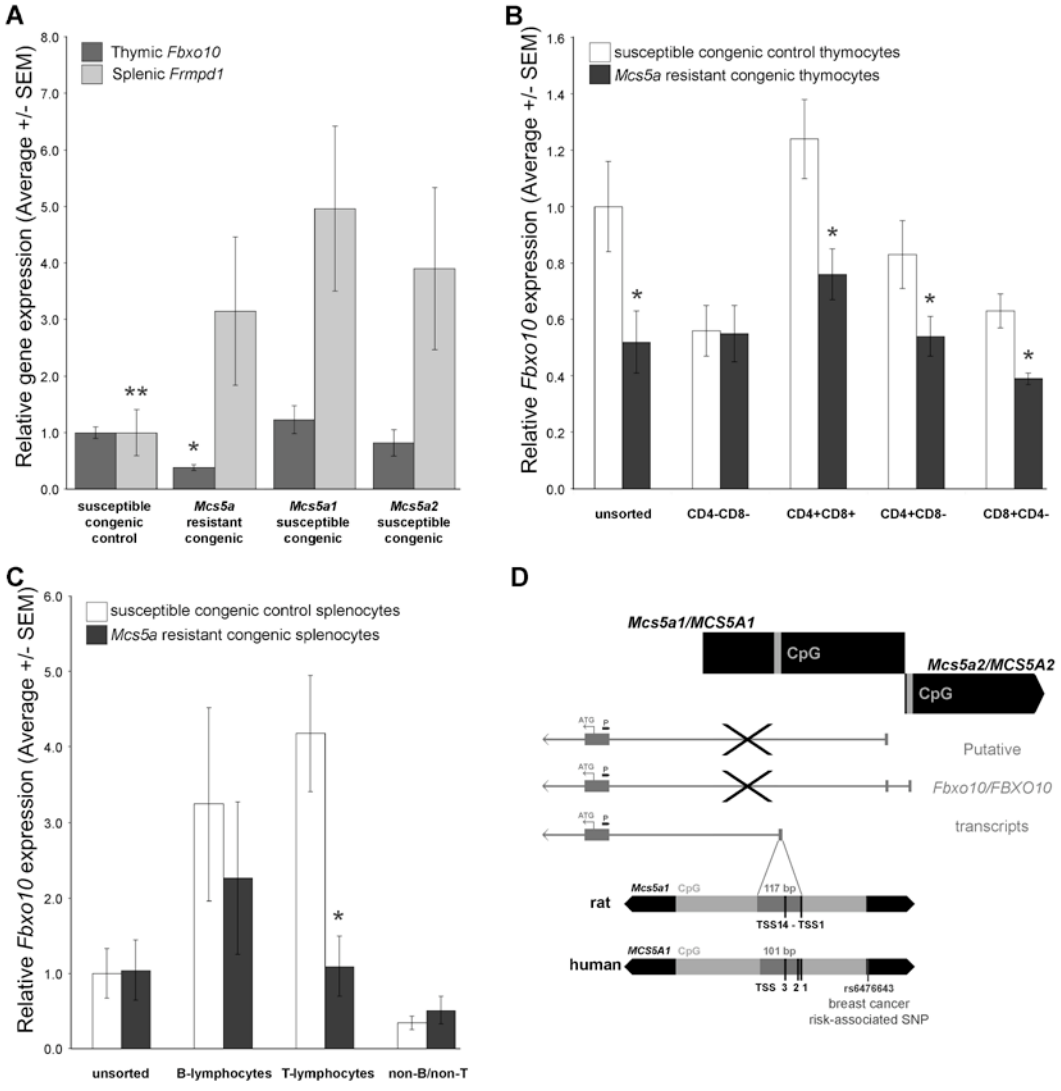

Figure 5: *Fbxo10* And *Frmpd1* Transcript Level Studies. **a)** Differential transcript levels of *Fbxo10,* but not *Frmpd1*, are related to the *Mcs5a*-associated mammary carcinoma resistance phenotype. Graphed are the average mRNA levels (normalized to *Gapdh*) ± SEM relative to the thymic *Fbxo10* (dark grey) and splenic *Frmpd1* (light grey) expression levels of the susceptible congenic control line. Susceptible congenic control line: n=9, *Mcs5a* congenic resistant line: n=6, *Mcs5a1* susceptible congenic line: n=9, and *Mcs5a2* susceptible congenic line: n=8. The *Mcs5a* resistant congenic line is the only line to have significantly lower

thymic *Fbxo10* transcript levels (one asterisk: P=0.01). Increased splenic *Frmpd1* transcript levels were detected in the *Mcs5a* resistant congenic, *Mcs5a1* susceptible congenic, and *Mcs5a2* susceptible congenic lines compared to the susceptible congenic control line (two asterisks: P=0.04). **b)** *Fbxo10* transcript levels in unsorted and flow cytometry-sorted thymocytes. Graphed are the average *Fbxo10* mRNA levels (normalized to *Gapdh*) ± SEM relative to the average transcript level of *Fbxo10* of the unsorted thymocytes of the susceptible congenic control line. The difference in *Fbxo10* transcript levels between the susceptible congenic control line and the *Mcs5a* resistant congenic line was identified in unsorted (P=0.03), CD4+CD8+ (P=0.02), CD4+CD8- (P=0.03), and CD8+CD4- (P=0.002) thymocytes (all significant differences are indicated with an asterisk). Sample sizes for the susceptible congenic control line and *Mcs5a* congenic resistant line were, respectively: unsorted n=16 and n=9, CD4-CD8- n=4 and n=4, CD4+CD8+ n=17 and n=19, CD4+CD8- n=15 and n=18, CD8+CD4- n=18 and n=17. **c)** *Fbxo10* transcript levels in unsorted and flow cytometry-sorted splenocytes. Graphed are the average *Fbxo10* mRNA levels (normalized to *Gapdh*) ± SEM relative to the transcript levels of *Fbxo10* for the unsorted splenocytes of the susceptible congenic control group. Differential *Fbxo10* transcript level was identified in sorted splenic CD3+ T-lymphocytes (indicated with an asterisk; P=0.002). Sample sizes for the susceptible congenic control line and *Mcs5a* resistant congenic line were, respectively: unsorted n=16 and n=12, B-lymphocytes n=9 and n=8, T-lymphocytes n=16 and n=13, non-B-/non-T-lymphocytes n=3 and n=5. d) *Fbxo10/FBXO10* Transcriptional Start Site analysis by the 5' RLM-RACE assay. Three putative *Fbxo10/FBXO10* transcripts are indicated relative to the position of the *Mcs5a/MCS5A* locus (in black). The location of the *Fbxo10/FBXO10* specific primers in the first coding exon is indicated (P). The 5' RLM-RACE revealed a cluster of 14 rat *Fbxo10* TSSs within the *Mcs5a1* CpG island (117 bp; indicated in light grey) and a cluster of 3 human *FBXO10* TSSs in the orthologous *MCS5A1* CpG island (101 bp, indicated in light grey). The human TSS cluster was found to be located at 150 bp distance from breast cancer risk-associated SNP rs6476643. The transcripts not identified in rat or human immune tissue are indicated with a X.

To understand the regulation of the *Fbxo10/FBXO10* gene it is important to know the exact transcriptional start site (TSS). This facilitates localization of the putative regulatory elements. Two areas of transcriptional initiation of the *Fbxo10/Fbxo10* gene in rats and humans are annotated, namely the CpG island in *Mcs5a1/MCS5A1* and an area close to the *Mcs5a1/MCS5A1-Mcs5a2/MCS5A2* border (Fig. 5d). To identify the TSSs of the *Fbxo10* gene, we performed the 5' RLM-RACE assay (RNA Ligase Mediated-Rapid Amplification of cDNA Ends). The assay makes use of an RNA-adapter that is ligated to the de-CAP-ped 5' end of transcripts, followed by a reverse transcriptase (RT) reaction to make cDNA. To amplify just the 5' ends of the *Fbxo10* gene, a nested PCR reaction was performed with primers annealing to the translational start codon (ATG)-containing exon and universal primers annealing to the 5' RNA-adapter. The assay was done on pools of thymus or spleen RNA from four susceptible congenic control and four *Mcs5a* resistant congenic rats, and on human RNA samples from thymus, spleen and breast tissue. A typical PCR product yielded multiple bands on an agarose gel, indicative of multiple TSSs. In total, 150 rat and 54 human TSS clones were sequenced to elucidate the exact start position of the *Fbxo10/FBXO10* transcripts. In the rat 14 TSS positions were found, all located in the CpG island of the *Mcs5a1* locus (Fig. 5d; CpG1in Fig. 6). In the human 3 TSS positions were found, again, located in the CpG island of the *MCS5A1* locus. It should be noted that in the human breast RNA sample the *FBXO10* TSS cluster was found to be located close to the *MCS5A1-MCS5A2* border (CpG2 in Fig. 6), indicating that the assay was able to identify such cases. Identification of the *Fbxo10/FBXO10* TSSs in the CpG island of *Mcs5a1/MCS5A1* suggests that potential regulatory elements (e.g. promoter or enhancers) may be located in the same region. The human TSS cluster is located within 150 bp of SNP rs6476643 (Fig. 5d), which is one of the breast cancer risk-associated SNPs found in *MCS5A1* (Samuelson et al., 2007).

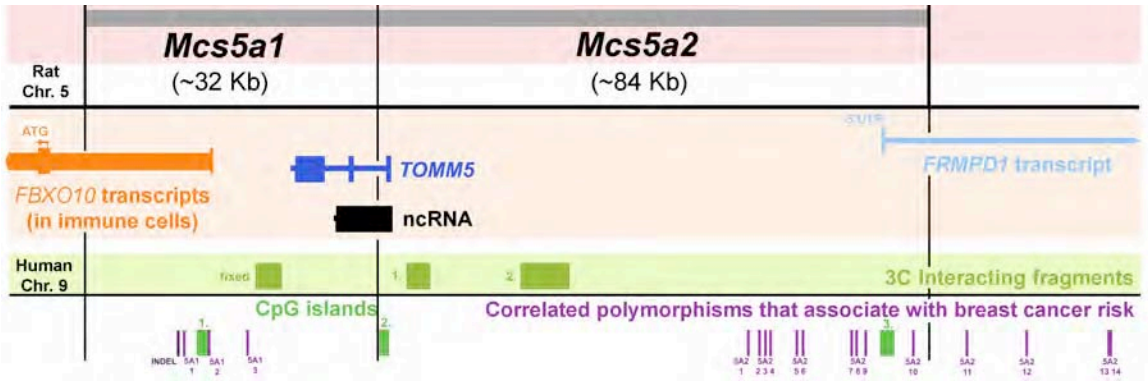

Figure 6: The *MCS5A* locus in a flat representation. Note that there are three major transcriptional start site areas associated with CpG islands (in dark green). The predominant *FBXO10* transcript in immune cells is displayed in orange. The *FRMPD1* transcript is shown in light blue. The correlated polymorphisms that associate with risk are shown as purple bars. The SNPs are numbered according to their position in *MCS5A1* (INDEL, 5A1_1-3), or *MCS5A2* (5A2_1-14). The interacting elements in the 3C assay are shown in light green. Other transcripts that start off from the promoter close to the *MCS5A1/ MCS5A2* border are shown in blue (*TOMM5*) and black (ncRNA).

Transcriptional activity of the breast cancer alleles (SoW Task 3 and 4)

The causative genetic variants of the breast cancer locus *MCS5A* are non-coding, suggesting a role in the regulation of gene transcription. Rat studies on gene expression regulation in various tissues identified the expression of the *Fbxo10* gene in T-lymphocytes to be associated with the mammary carcinogenesis phenotype. Therefore, we hypothesize that the breast cancer alleles regulate the expression of the *Fbxo10* gene in T-lymphocytes.

To mechanistically study how the correlated polymorphisms regulate transcription, I employed the pGL3-luciferase (LUC) expression vector system (Promega) in a human T-lymphocytic cell culture model system (JURKAT). The system was successfully established by transiently transfecting JURKAT cells with control vectors having known luciferase activity (pGL3-basic, and pGL3-control) together with a vector resulting in renilla (REN) activity for internal normalization (SoW Task 3a). LUC and REN activities are read using a luminometer. Transcriptional activity is calculated as the ratio LUC/REN normalized against the activity of the control vector transfected in the same experiment.

The higher-order chromatin interactions between *Mcs5a1* and *Mcs5a2* found by the 3C assay are not allele dependent. Therefore, they are not solely responsible for the expression regulation of the *Fbxo10* gene in the T-lymphocytes and ultimately breast cancer susceptibility. Subsequently, screening the interacting elements for transcriptional activity (as outlined in SoW Task 3b) would not explain how the *Fbxo10* gene is regulated. The 5' RLM-RACE assay indicated the start site and putative proximal promoter of the *FBXO10* gene in the human to be located amidst the correlated polymorphisms associated with breast cancer risk in *MCS5A1*. Hence, I decided to first screen all breast cancer-associated polymorphisms (both alleles, if available) for promoter activity (Fig. 7). Other (promoter) elements in the locus were also included to establish base line promoter activity levels. This revealed that there are three main areas of promoter activity in the region, closely associated with the predicted transcripts. The activity of the *TOMM5* transcript promoter is strongest. Since the expression level of the *Fbxo10* is correlated with the mammary tumor multiplicity phenotype, I decided to focus

on the *FBXO10* promoter and the 4 correlated breast cancer polymorphisms surrounding it. Only the fragment containing SNP 5A1_2 (rs6476643 in Fig. 5d) showed promoter activity (Fig. 7), which is in accordance with the identified start sites of the *FBXO10* gene. The promoter activity is not different upon introduction of the susceptible allele (Fig. 7). This is not surprising as the rat expression level study suggested that the resistant allele of both *Mcs5a1* and *Mcs5a2* needs to be present to reduce *Fbxo10* expression.

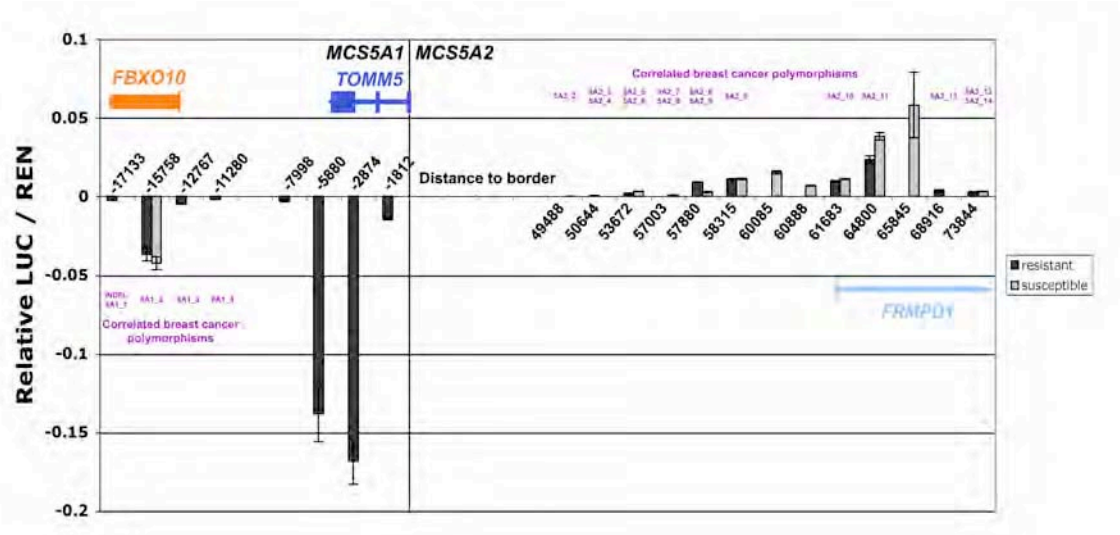

Figure 7: Promoter activity in the *MCS5A* locus. Each element screened for promoter activity in the luciferase assay is between 800-1,400 bp in size. The number indicating each element represents the genomic distance of the middle of the element from the *MCS5A1-MCS5A2* border. The predominant *FBXO10* transcript in immune cells is displayed in orange. The *FRMPD1* transcript is shown in light blue. The *TOMM5* transcript is indicated in dark blue. The distribution over the screened fragments of the correlated polymorphisms that associate with breast cancer risk is shown in purple.

It should also be mentioned that some elements containing breast cancer-associated SNPs in *MCS5A2* show promoter activity, of which one even shows differential promoter activity (SNP 5A2_11; Fig. 7). These elements, however, are associated with the promoter of the *FRMPD1* gene, whose expression in the rat is not correlated with the mammary carcinogenesis susceptibility phenotype. Thus, these elements may not contribute to breast cancer susceptibility as promoter elements, but may have a distal effect on the expression level of the *FBXO10* gene.

To test the hypothesis that the polymorphisms of the resistant risk-associated alleles of *MCS5A1* and *MCS5A2* could interact to downregulate *FBXO10* transcript levels in human T-lymphocytes, Luciferase reporter assays were performed using a series of constructs containing selected *MCS5A1* and *MCS5A2* alleles. All constructs described below were visually inspected by restriction enzyme digest to ensure integrity (Fig. 8). Again, the human T-lymphocytic cell line (JURKAT) was transiently transfected with each construct in the presence of a REN expressing vector as a control for transfection efficiency.

First, a 1,464 bp fragment of *MCS5A1* including the previously identified *FBXO10* TSS cluster and SNP rs6476643 (SNP 5A1_2 in Fig. 7) was inserted upstream of the *LUC* reporter gene in the pGL3-basic vector (Fig. 9a). Two versions of the construct were made, one containing the resistant (R) allele of rs6476643 (G) and one containing the susceptible (S) allele of rs6476643 (T). The transcriptional activity of the R and S

version of the *FBXO10* TSS-Luciferase construct is not significantly different (P=0.45; Fig. 9b).

Subsequently, the R and S *FBXO10* TSS-Luciferase constructs were modified to harbor *MCS5A2* fragments containing the resistant or susceptible allele of one or two *MCS5A2* SNPs (Fig. 9c). The *MCS5A2* fragments were inserted downstream of *Luc+* to mimic the long-range nature of the putative interaction. The 15 correlated *MCS5A2* polymorphisms were integrated into 10 *MCS5A2* fragments of on average ~990 bp in length (Table 1). This yielded a series of 10 constructs, each present in two versions, namely SNP rs6476643 susceptible plus *MCS5A2* susceptible (SS) and SNP rs6476643 resistant plus *MCS5A2* resistant (RR). The transcriptional activity of the entire series of 10 RR constructs is significantly lower compared to the entire series of 10 SS combination constructs (Wilcoxon rank test $P<10^{-48}$).

Next, the transcriptional activity of the SS and RR versions was compared for each construct (Fig. 9d). The transcriptional activity was not statistically significantly different between the SS and RR versions of constructs 1, 2, 5, 7, 8 (P>0.05, Table 1). However, the activity of constructs 3, 4, 6, 9, and 10 were significantly lower in the RR version compared to the SS version (P<0.05, Table 1). In addition, the *MCS5A2* SNPs rs62534439 (construct 4), and rs62534443 and rs62534444 (constructs 5, 6) were also represented in constructs 4a (Fig. 9e) and 5a (Fig. 9f), respectively, again resulting in lower activities for the RR versions compared to the SS versions, although only significant for construct 4a (P<0.05, Table 1).

These findings implicate that a fragment containing *MCS5A1* SNP rs6476643 and the *FBXO10* TSSs is not displaying altered transcriptional activity between the R and S allele. However, when combined with the corresponding allele of the *MCS5A2* SNPs, the transcriptional activity of the RR versions of the constructs is significantly lower compared to the SS versions. This is consistent with the observation of lower *Fbxo10* expression in the T-lymphocytes of *Mcs5a* resistant congenic animals compared to susceptible congenic control animals.

Table 1: Characteristics and statistical results of Luciferase assays.

Fragment*	**Coordinates**	**Size (in bp)**	**dbSNP ID**	**P-value****
FBXO10 TSS	chr9:37,575,150-37,576,613	1,464	rs6476643	0.4492
MCS5A2 fragment 1	chr9:37,629,700-37,630,581	882	rs4878708, rs4878709	0.8182
MCS5A2 fragment 2	chr9:37,630,824-37,631,694	871	rs4878710	0.937
MCS5A2 fragment 3	chr9:37,631,683-37,632,895	1,213	rs2182317, rs10973450	0.000329
MCS5A2 fragment 4	chr9:37,634,580-37,635,496	917	rs62534439, rs3075980	0.002165
MCS5A2 fragment 5	chr9:37,638,084-37,639,200	1,117	rs4490927, rs62534443	0.4359
MCS5A2 fragment 6	chr9:37,639,500-37,640,408	909	rs62534444	0.001088
MCS5A2 fragment 7	chr9:37,642,903-37,643,741	839	rs62534445	0.1949
MCS5A2 fragment 8	chr9:37,645,852-37,647,026	1,175	rs4878713	0.06494
MCS5A2 fragment 9	chr9:37,650,036-37,651,074	1,039	rs55677371	0.008658
MCS5A2 fragment 10	chr9:37,655,040-37,655,926	887	rs62534456, rs62534457	0.02597
MCS5A2 fragment 4a	chr9:37,633,876-37,634,898	1,023	rs62534439	0.04113
MCS5A2 fragment 5a	chr9:37,639,062-37,639,976	915	rs62534443, rs62534444	0.08298

* The *FBXO10* TSS represents a fragment encompassing the TSS of the *FBXO10* transcripts as determined in the 5'RLM-RACE assay. The *FBXO10* TSS (harboring the S or R allele of rs6476643) is present in all constructs. The fragments containing the *MCS5A2* breast cancer risk-associated variants were inserted into a cloning site downstream of the *Luc+* reporter gene.

** P-value in the Wilcoxon rank sum test, comparing the Luciferase activity of the SS and RR versions of the construct.

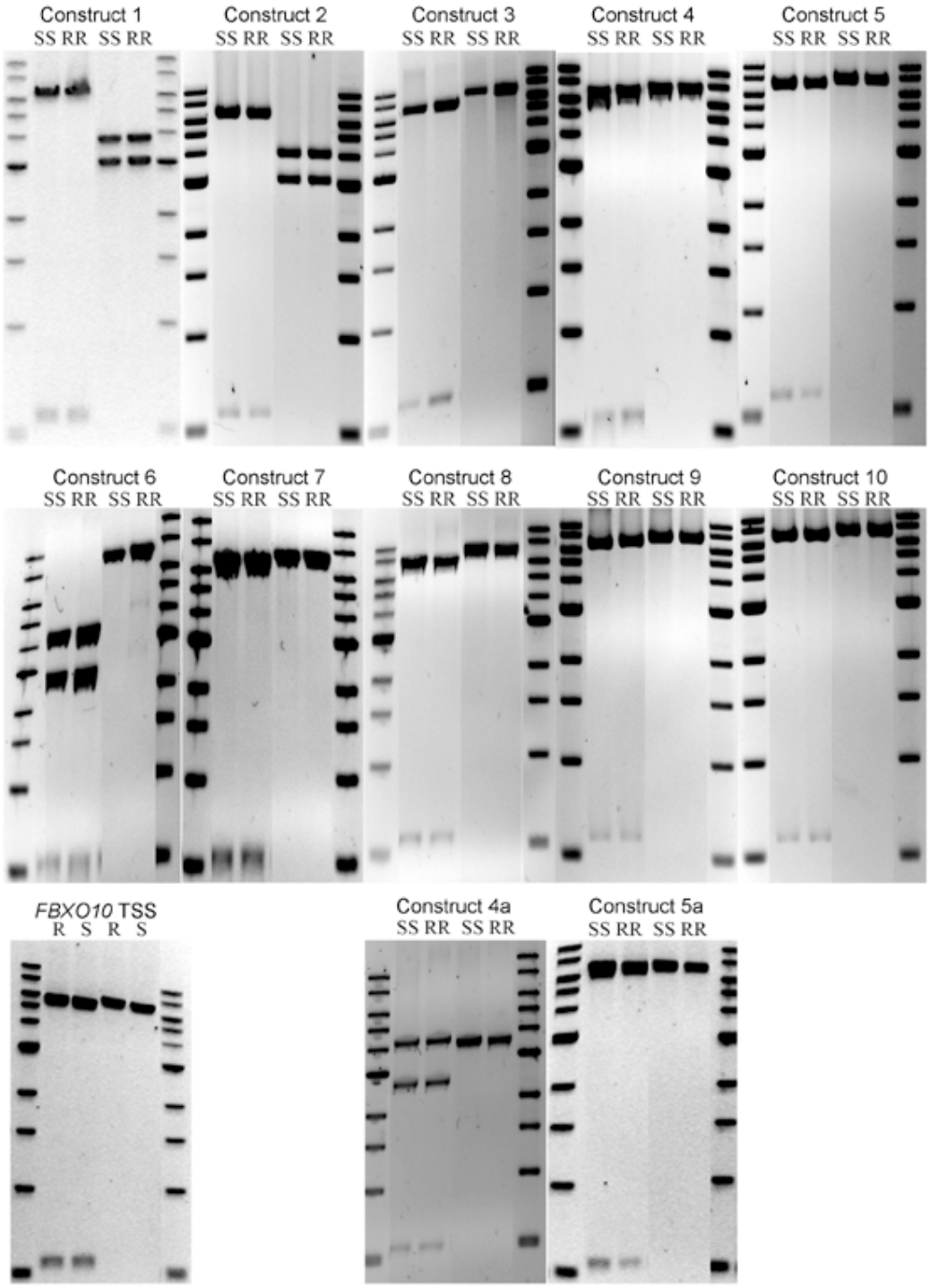

Figure 8: Visual Inspection of the Integrity of the Luciferase Constructs. Between 100 and 200 ng of each vector was digested with two restriction enzymes, namely *Not*I and *Sac*I and analyzed using agarose gel electrophoresis. The DNA ladder used in all gels is the 1 Kb ladder. RR=R allele of *MCS5A1* SNP rs6476643 combined with the R allele of SNP(s) present in the *MCS5A2* fragment. SS=S allele of *MCS5A1* SNP rs6476643 combined with the S allele of SNP(s) present in the *MCS5A2* fragment.

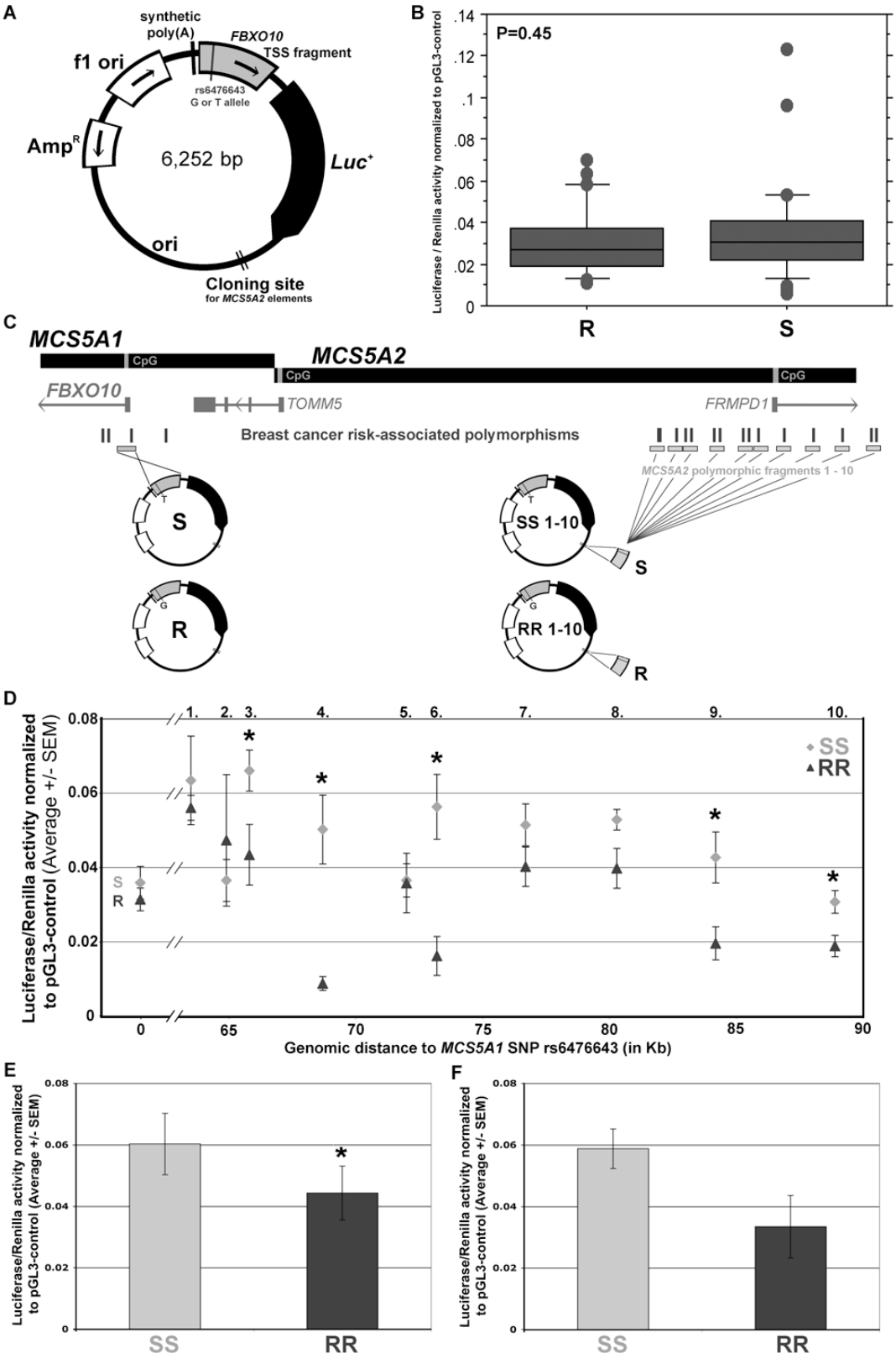
A
synthetic poly(A)
FBXO10 TSS fragment
f1 ori
rs6476643 G or T allele
Amp^R
6,252 bp
Luc+
ori
Cloning site for MCS5A2 elements
B
P=0.45
Luciferase / Renilla activity normalized to pGL3-control
R
S
C
MCS5A1
MCS5A2
CpG
FBXO10
TOMM5
FRMPD1
Breast cancer risk-associated polymorphisms
MCS5A2 polymorphic fragments 1 - 10
S
R
SS 1-10
RR 1-10
D
Luciferase/Renilla activity normalized to pGL3-control (Average +/- SEM)
SS
RR
Genomic distance to MCS5A1 SNP rs6476643 (in Kb)
E
F
SS
RR

Figure 9: Transcriptional Activity Analysis of the Human Breast Cancer Risk-associated Susceptible and Resistant Alleles in Luciferase Reporter Assays. **a)** Map of the *FBXO10* TSS-Luciferase reporter construct. A *MCS5A1* fragment containing the *FBXO10* TSSs and the risk-associated SNP rs6476643 was inserted upstream of the Luciferase reporter gene (*Luc*$^{+}$) in reverse genomic orientation. Two versions of the construct were created, namely having the susceptible (S; T allele) or resistant (R; G allele) of SNP rs6476643. Other features of the construct include the ampicillin resistance gene (Amp^{R}), origin of replication derived from filamentous phage (f1 ori), origin of replication in *E. coli* (ori), a synthetic poly(A) signal/transcriptional pause site for background reduction (synthetic poly(A)) and a cloning site downstream of *Luc*$^{+}$. Arrows indicate direction of transcriptional activity. **b)** Boxplot of the relative Luciferase activity of the R and S constructs. n=30 transient transfection assays. **c)** Schematic representation of the position of the genomic fragments derived from the *MCS5A1* and *MCS5A2* loci combined into the Luciferase constructs. The *MCS5A1* and *MCS5A2* loci are shown as black lines. The light grey bars within the black lines represent the CpG islands located in the locus. The genes are shown in dark grey. The breast cancer risk-associated polymorphisms are represented as vertical grey lines. The fragments subcloned into the reporter constructs are indicated as horizontal light grey bars. The susceptible alleles of the *MCS5A2* polymorphisms were combined with the susceptible allele of the *MCS5A1* SNP rs6476643 (SS constructs 1-10). Similarly, the resistant alleles were combined (RR constructs 1-10). **d)** Relative Luciferase activity of constructs SS 1-10 and RR 1-10 versus the genomic distance of the *MCS5A2* polymophisms to the *MCS5A1* SNP rs6476643. The values at genomic distance 0 correspond to the *FBXO10* TSS-Luciferase constructs S and R. The measurements indicated with an asterisk are significantly different between SS and RR (P<0.05). n=6 or more transient transfection assays. **e)** Relative Luciferase activity of constructs SS 4a and RR 4a. The SS and RR values are significantly different (indicated with an asterisk; P<0.05). **f)** Relative Luciferase activity of constructs SS 5a and RR 5a.

Elucidating the mechanism of the breast cancer alleles (SoW Task 5)

The final stage of the project focuses on elucidating the transcription factors that differentially bind to the breast cancer SNPs that are implicated in regulating the expression of the *FBXO10* gene. Using TFSEARCH I performed computational transcription factor binding site predictions on the SNPs present in the fragments that showed reduced transcriptional activity in the Luciferase assay (SoW Task 5a).

```
rs6476643       GGGCTGGGCTTCCCGACCACCGCGCA[G/T]AAAAGCTGTATCTGCAGGAGGGGCA
MCS5A1                   cdx-a to both alelles, TATA to T-allele

rs2182317       AACAGAAGCCCCTTGTAGAGTACAGG[A/C]ATAAGCAGAGTAAATCTAAATGAAA
                        No vertebrate binding sites

rs10973450      GGTTACTTAACCATGTAGAGTCTCTT[C/T]GTCTGCAAAAAAGACACATGATACT
                        No vertebrate binding sites

rs62534439      CCAGCTCTGTGACCTCATACCAGTCG[C/T]TTGAATTCTCTGAGCTTGCCTCAGT
                        No vertebrate binding sites

rs62534444      ATGCACTTGTTAACATCTGCCTGTGC[A/G]CCATCCCCAGAATGATCTAACATCC
                        Gata-1 and Gata-2 to both alleles

rs55677371      AGTGATTTTAAAGTAGGTTTAAACAA[C/T]GGGTTTAAAGAACAGTGATTTTCCA
                        v-Myb to C-allele       Sox-5 to T-allele
                        SRY to C-allele         SRY to T-allele

rs62534456      TATGTTGAAAATGTGTCTTTTCACAC[A/T]AAAAGACTGGAAGAGTAATTAGCAA
                        C/EBP to A-allele

rs62534457      CTTGAACTTCTGAAATTATTTTTTCC[A/G]CTCCATTTGTAATTGAGCCCAGGGA
                         c-Ets to G-allele
```

Electromobility shift assays (EMSA) and chromatin immunoprecipitations (ChIP) could be initiated to reveal differential binding of the predicted factors to the alleles (SoW Task 5b-d). In the mean time, I decided to focus on the effect of the locus on global gene expression regulation. To identify transcriptome changes underlying potential alterations in T-lymphocyte function between susceptible congenic control and *Mcs5a* resistant congenic animals, a global gene expression study was performed. Total RNA extracted from T-lymphocyte-enriched splenocytes of nine susceptible congenic control and eight *Mcs5a* resistant congenic animals was mixed into three pools per genotype group (Figure 5A). The six RNA pools were profiled by DGE, a next-generation sequencing approach evolved from Massively Parallel Signature Sequencing (MPSS) technology (Brenner et al., 2000; Jongeneel et al., 2003). DGE relies on next-generation sequencing of 17 bp signature tags representing specific mRNA molecules (t Hoen et al., 2008). Per RNA pool this procedure produced roughly six to seven million 17 bp tags. A global transcriptional profile for each RNA pool was constructed by mapping the tags to the rat genome and assigning the tags and their counts to the transcriptome. Only the count for the most abundant, sense tag closest to the 3' end of a transcript was included, leaving >2M counts per sample (Fig.11a). A total of 15,834 genes were classified as 'non-expressed' (NONEX), as these genes obtained less than seven counts across the six RNA pools. A total of 11,354 genes were classified as 'expressed' (EX). Amongst the 'expressed' class 198 genes were found to be DE (EX DE, adjusted $P<0.05$, Table 2) and 11,156 genes non-DE (EX NONDE). The EX DE genes were divided into 95 genes upregulated in the *Mcs5a* resistant congenic rats (EX DE *Mcs5a*-over) and 103 genes downregulated in the *Mcs5a* resistant congenic animals (EX DE *Mcs5a*-under).

The expression regulation of *Fbxo10* suggests a repressive activity of the resistant *Mcs5a* allele, however, gene expression regulation cannot be solely regarded as a local event. To validate the role of *Mcs5a* as a transcriptional repressor, we investigated the nuclear environment in which *Mcs5a* resides. As it is unclear which transcription factors are involved in the regulatory activity of *Mcs5a*, the 'open-ended' circular chromosome conformation capture (4C) technology in combination with next-generation sequencing was implemented using the fixed fragment in *Mcs5a1* that showed looping to *Mcs5a2* and binding of CTCF/cohesin as the bait. Two pools of 3C templates from six heterozygous *Mcs5a* resistant congenic animals (*cis* het) and six heterozygous animals from an intercross between both *Mcs5a1* and *Mcs5a2* susceptible congenic animals (*trans* het), were transformed into 4C libraries, as previously described (Simonis et al., 2006). Briefly, the 3C template pools were digested with a 4 bp cutter, *Nla*III (Fig.10). To circularize the molecules, a ligation was done in strongly dilute fashion. Captured elements were amplified in a linear PCR, which yielded a smear of DNA fragments representing a potential collection of fragments originally ligated to the fixed *Bgl*II fragment in *Mcs5a1* in the 3C assay. The products were analyzed using a next-generation sequencing approach. The reads were mapped to the rat genome; however, only reads that mapped between a *Bgl*II site and a *Nla*III site were included in the analysis. This yielded ~1,7M and ~1,1M reads for the '*cis* het' and '*trans* het' library, respectively (Fig. 11a). For each *Bgl*II fragment, the expected amount of reads was calculated. Using a negative binomial statistical model, 10,044 *Bgl*II fragments having a significantly more than expected number of reads were categorized as the '*Mcs5a*-associated' *Bgl*II fragments. Of these, 2,637 *Bgl*II fragments displayed significantly different amount of reads between the '*cis* het' and '*trans* het' libraries (Fig. 11a).

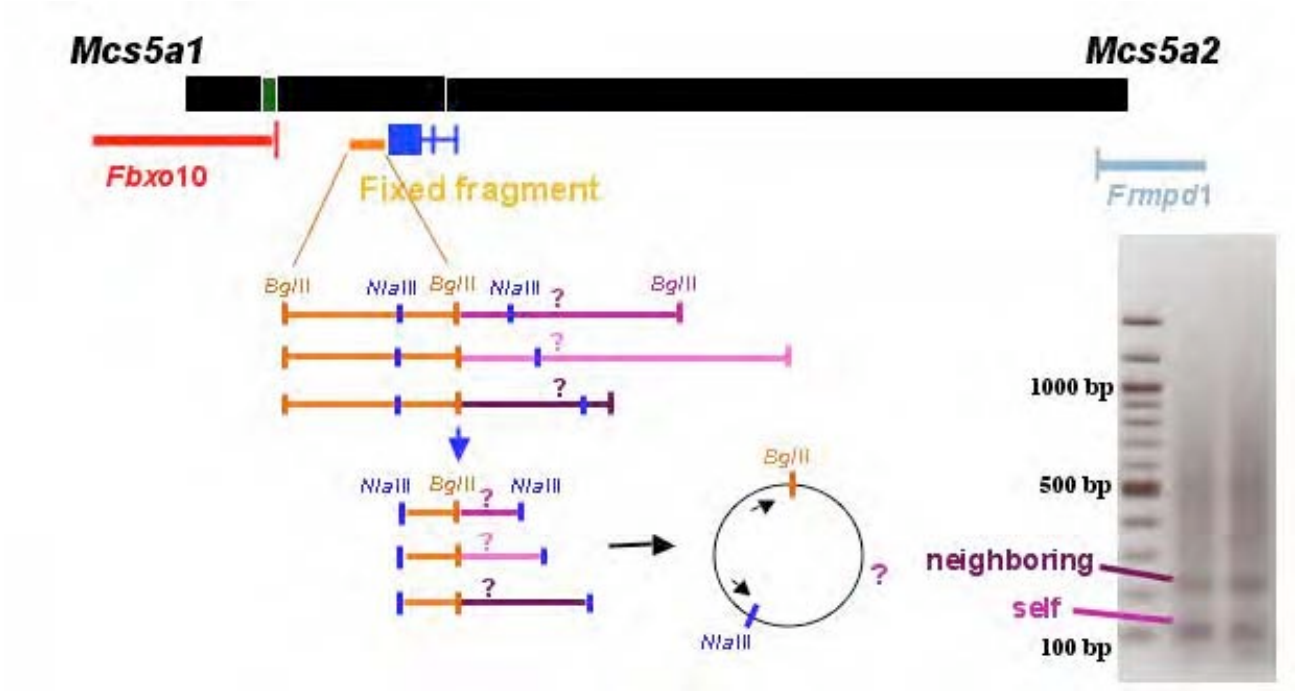

Figure 10: Schematic Representation of the 4C Approach. As the bait fragment for the 4C assay (shown as an orange bar), the fixed fragment in *Mcs5a1* that showed looping to the *Mcs5a2* locus in the 3C assay was chosen. Two pools of 3C samples of splenic T-lymphocytes of six heterozygous *Mcs5a* resistant congenic (*cis* heterozygous) and six heterozygous *Mcs5a1* x *Mcs5a2* susceptible congenic (*trans* heterozygous) animals were converted into two 4C libraries. Therefore, the 3C libraries were digested with *Nla*III, a 4 bp cutter. The resulting molecules were circularized by ligation, and the captured fragments were amplified in an inverse PCR, using two primers on the bait fragment directed towards the *Bgl*II and *Nla*III sites. The amplified captured fragments from both libraries (*cis* and *trans*) were run on an agarose gel. The amplification produced a smear of putative *Mcs5a*-interacting fragments. The visible bands in the two smears represent the signal for the self-ligating *Bgl*II bait fragment (110 bp), and the neighboring *Bgl*II fragment as a result of incomplete digestion (220 bp). To identify the relative abundance of the other fragments in the smear, a next-generation sequencing approach was employed.

To verify some of the interchromosomal interactions, fluorescent in-situ hybridization (FISH) was carried out using nick-translated, fluorescently labeled BAC probes on interphase nuclei of splenic T-lymphocyte-enriched samples of susceptible congenic control and *Mcs5a* resistant congenic rats. Six of the *Mcs5a*-associated fragments and one unassociated control fragment were selected. FISH interactions between the selected fragments and *Mcs5a* were scored if one pair of alleles colocalized (Table 3). Additionally, for each selected fragment the tag density to the entire BAC region was calculated from the 4C data (Table 3). The two BACs having the highest tag density in the 4C assay showed the highest FISH interaction frequency with the *Mcs5a* locus (Fig. 11b). The four BACs with intermediate tag density in 4C showed intermediate FISH interaction frequencies (Fig. 11b). Finally, the unassociated control BAC showed the lowest FISH interaction frequency. The FISH interactions recapitulated the 4C results for the selected fragments.

The 4C results suggest that *Mcs5a* colocalizes in the nuclei of a T-lymphocyte population with thousands of chromosomal regions. The distribution of the *Mcs5a*-associated regions over the rat chromosomes is non-random (Fig. 11c). Chromosomes *5*, *8*, and *12* harbor significantly more than expected *Mcs5a*-associated fragments (adjusted $P<0.05$), whereas chromosomes *2*, and *16* harbor significantly less than

expected *Mcs5a*-associated fragments (adjusted P<0.05). This observation likely reflects the organization of the nucleus in chromosomal territories (Lieberman-Aiden et al., 2009). Chromosome *5* showed the greatest degree of deviation from the expected number of *Mcs5a*-associated fragments (adjusted $P<10^{-77}$). This is expected, as *Mcs5a* is residing on chromosome *5*, and intrachromosomal interactions are more likely to form than interchromosomal interactions (Lieberman-Aiden et al., 2009).

For each *Bgl*II fragment the genomic distance to the nearest gene was calculated. For groups of *Bgl*II fragments with decreasing tag densities the average distance to the nearest gene was determined. For chromosome *5*, as well as for all other chromosomes, it was found that groups of *Bgl*II fragments with the highest tag densities have the lowest average distance to the nearest gene (Fig. 11d).

Next, we asked if the *Mcs5a*-associated fragments relate to the global gene expression profile of the T-lymphocytes. Therefore, for each gene region (the gene span plus 10 Kb of upstream sequence) was calculated if it overlapped with a *Mcs5a*-associated fragment. For each gene category (determined by the DGE analysis) the expected hit rate (gene regions overlapping with *Mcs5a*-associated fragments) and non-hit rate (gene regions not overlapping with *Mcs5a*-associated fragments) were calculated, taking into account gene region sizes. The observed hit rate / non-hit rate distribution was tested with the expected hit rate / non-hit rate distribution in a chi-square test for a 2x2 contingency table (Table 4). When comparing EX to NONEX genes, the observed distribution deviated highly significantly ($P<10^{-152}$) from the expected distribution, in favor of the EX category (Fig. 11e). When comparing EX DE to EX NONDE genes within the EX category, the observed distribution does not significantly deviate from the expected (p=0.08), although there is a trend towards the EX DE category having more *Mcs5a*-associated fragments overlapping (Fig. 11f). However, when the *Mcs5a*-over and *Mcs5a*-under expressed genes within the EX DE category were compared, the observed distribution did significantly differ from the expected (p=0.01), with the *Mcs5a*-under expressed category having more *Mcs5a*-associated fragments overlapping. Finally, when only *Mcs5a*-associated fragments that are different between the '*cis* het' and '*trans* het' libraries were included in the hit rate / non-hit rate calculations of the EX DE and EX NONDE categories, the observed distribution did not deviate from the expected (p=0.48; Table 4).

In summary, *Mcs5a* preferably colocalizes with EX DE *Mcs5a*-under expressed gene regions in the nuclei of a population of T-lymphocytes. This finding substantiates the repressive gene regulatory role of the resistant *Mcs5a* allele, which is in accordance with the previous finding of *Fbxo10* transcript level downregulation.

Figure 11: Results of the Circular Chromosome Conformation Capture (4C) Assay and the Digital Gene Expression Assay (DGE). **a)** Overview of the datasets used in the DGE and 4C studies. The left portion of the panel reports the results of the DGE assay, the right portion reports the results of the 4C assay. Total RNA was mixed into three pools for the susceptible congenic control genotype (S1-3) and three pools for the *Mcs5a* resistant congenic genotype (R1-3). 3C samples from *Mcs5a* resistant congenic (*cis*) heterozygous animals and *Mcs5a1*x*Mcs5a2* susceptible congenic (*trans*) heterozygous animals were mixed into two pools that were transformed into 4C libraries '*cis* het' and '*trans* het'. Statistical approaches identified *Mcs5a*-associated *Bgl*II fragments, of which a subset was found to be differentially hit in the '*cis* het' and '*trans* het' libraries (Diff). **b)** Fluorescent In-Situ Hybridization (FISH) with Bacterial Artificial Chromosome (BAC) probes confirmation study of seven regions with varying tag density in the 4C assay. The 4C tag density for a given BAC region was calculated as the sum of the normalized tag counts of the *Bgl*II restriction fragments localizing to the mappable portion of the region, divided by the length (in bp) of the mappable portion of the region (tpm/bp). The BAC interaction frequency is calculated as the frequency of occurrence of an overlap between a variable BAC signal with one of the two *Mcs5a* BAC signals. **c)** Distribution of the *Mcs5a*-associated *Bgl*II fragments across the chromosomes. The number of *Mcs5a*-associated fragments is expressed as a fraction of the expected number of *Mcs5a*-associated fragments for each chromosome (set to 1) if the *Mcs5a*-associated fragments were distributed randomly. Chromosomes with a significantly

different relative amount of *Mcs5a*-associated fragments than expected are indicated with an asterisk (chi-square, Bonferroni adjusted P<0.05). **d)** The average distance to the nearest gene is increasing for groups of *Bgl*II fragments with decreasing tag density. The minimum tag density of a group of *Bgl*II fragments is plotted versus the average +/- SEM distance to the nearest gene (in bp) of a group of *Bgl*II fragments. **e)** Hit rate of NONEX and EX genes with *Mcs5a*-associated fragments. The observed hit rate / non-hit rate distribution deviated significantly from the expected (indicated with asterisks, Chi-square 2x2 contingency table P<0.05). **f)** Hit rate of EX NONDE and EX DE genes with *Mcs5a*-associated fragments. **g)** Hit rate of EX DE *Mcs5a*-over and EX DE *Mcs5a*-under genes with *Mcs5a*-associated fragments. The hit rate of a group of genes is calculated as the frequency of overlap between gene regions and *Mcs5a*-associated *Bgl*II fragments, corrected for the gene region size of the group. The expected hit rate is calculated as the hit rate if the *Mcs5a*-associated fragments were distributed uniformly over the gene regions, and is set to 1.

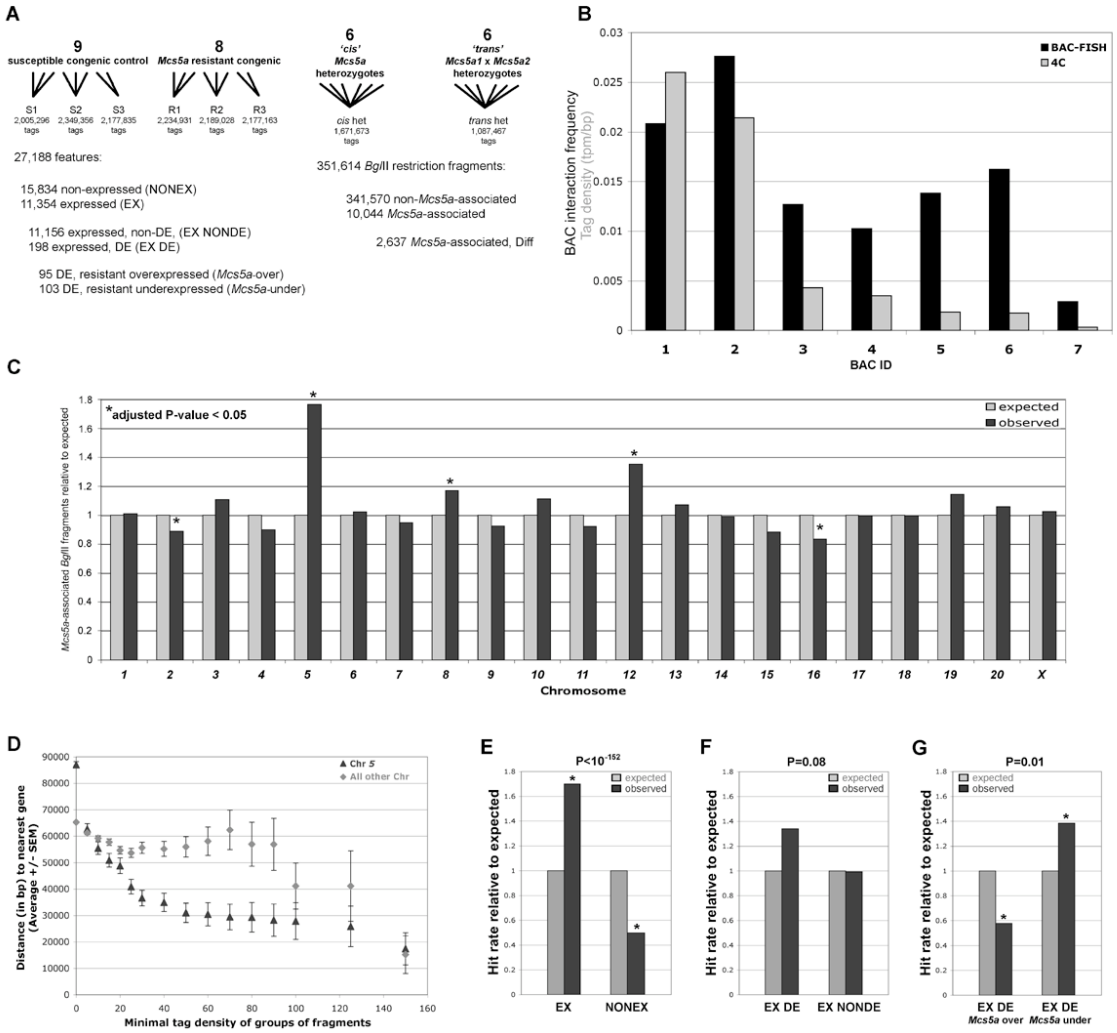

Table 2: List of 198 DE genes identified by Digital Gene Expression.

Mcs5a-over expressed genes				*Mcs5a*-under expressed genes			
Ensembl	**Name**	**P-value**	**S/R FC***	**Ensembl**	**Name**	**P-value**	**S/R FC***
ENSRNOG00000000111	RGD1565675_predicted	2.32E-34	0.6591	ENSRNOG00000003358	Ier5	2.64E-17	1.9734
ENSRNOG00000005538	Psmd11	2.24E-25	0.6703	ENSRNOG00000036695	Mrpl12	2.34E-16	1.6655
ENSRNOG00000019977	Ptprf	1.44E-16	0.6840	ENSRNOG00000005765	Ap4s1	5.23E-11	1.8667
ENSRNOG00000012803	Wdr61	3.84E-15	0.6616	ENSRNOG00000008734	Zmym2	6.60E-10	1.5578
ENSRNOG00000037904	Taf12	5.38E-13	0.7560	ENSRNOG00000021175	NP_001099802.1	8.99E-07	1.3581
ENSRNOG00000033217	Esam	2.16E-12	0.6055	ENSRNOG00000001005	Fcer2a	9.37E-07	3.2469
ENSRNOG00000026814	Qtrtd1	1.40E-10	0.6324	ENSRNOG00000017596	ENSRNOG00000017596	2.22E-06	1.2675
ENSRNOG00000019113	Hnrpk	8.17E-10	0.8163	ENSRNOG00000016952	LOC686442	2.92E-06	1.2988
ENSRNOG00000040303	1110001A16Rik	1.11E-09	0.6505	ENSRNOG00000014886	LOC498796	4.29E-06	1.2056
ENSRNOG00000019282	Ubqln1	1.19E-09	0.6632	ENSRNOG00000000926	NP_001099393.1	1.95E-05	1.7869
ENSRNOG00000010512	Yipf1	6.60E-09	0.7592	ENSRNOG00000001139	RGD1305986_predicted	2.42E-05	1.4337
ENSRNOG00000014999	Tnpo1	1.80E-08	0.7019	ENSRNOG00000007687	Sema7a_predicted	2.92E-05	1.9503
ENSRNOG00000037432	NP_001101443.1	2.33E-08	0.6118	ENSRNOG00000003150	RGD1563422_predicted	9.42E-05	1.3199
ENSRNOG00000016085	NP_001100288.1	3.58E-08	0.4115	ENSRNOG00000012640	Dpp7	0.0001	1.2376
ENSRNOG00000018637	Srcap	5.33E-08	0.7188	ENSRNOG00000015894	LOC499337	0.0001	1.1674
ENSRNOG00000029068	Wdr44	1.79E-06	0.6935	ENSRNOG00000019491	Stard10	0.0002	1.5818
ENSRNOG00000009397	NP_001101216.1	2.92E-06	0.7496	ENSRNOG00000004711	Mta1	0.0002	1.2435
ENSRNOG00000000692	Ung	3.32E-06	0.8032	ENSRNOG00000020436	RGD1311703	0.0002	1.2280
ENSRNOG00000012786	Pgrmc1	4.73E-06	0.4320	ENSRNOG00000017496	Cnp1	0.0002	1.2044
ENSRNOG00000019047	RGD1307982	5.67E-06	0.7887	ENSRNOG00000012379	Wdr18	0.0003	1.4143
ENSRNOG00000015616	Rgs14	9.46E-06	0.8344	ENSRNOG00000016734	Emilin3_predicted	0.0004	#DIV/0!
ENSRNOG00000014050	NP_001101302.1	1.46E-05	0.7992	ENSRNOG00000025443	Map1lc3a	0.0005	1.4926
ENSRNOG00000000108	Aga	2.13E-05	0.7847	ENSRNOG00000019859	Lypla3	0.0005	1.2956
ENSRNOG00000001039	Eif2b1	2.42E-05	0.4325	ENSRNOG00000026528	Mrps33_predicted	0.0005	1.2349
ENSRNOG00000032258	NP_001099430.1	3.08E-05	0.7548	ENSRNOG00000007356	Vps24	0.0006	1.2580
ENSRNOG00000002394	Tpr	7.91E-05	0.4670	ENSRNOG00000016280	Btrc	0.0007	1.9130
ENSRNOG00000023023	NP_001099801.1	0.0001	0.4963	ENSRNOG00000013370	Gfer	0.0007	1.5265
ENSRNOG00000018567	Slc20a1	0.0001	0.8147	ENSRNOG00000012266	NP_001102737.1	0.0009	1.8274
ENSRNOG00000011756	NP_001102261.1	0.0002	0.7457	ENSRNOG00000012937	NP_001099654.1	0.0009	1.3277
ENSRNOG00000029679	LOC361937	0.0002	0.4017	ENSRNOG00000012172	Sfpi1	0.0009	1.8212
ENSRNOG00000024503	4933425L03Rik	0.0002	0.5586	ENSRNOG00000032531	ENSRNOG00000032531	0.0010	3.4946
ENSRNOG00000026420	Iqcg	0.0005	0.5932	ENSRNOG00000029881	LOC688951	0.0010	1.9242
ENSRNOG00000008258	Fnbp1	0.0005	0.6832	ENSRNOG00000010664	NP_001100995.1	0.0014	2.1523
ENSRNOG00000000380	NP_001101097.1	0.0006	0.8193	ENSRNOG00000007104	Itpr1	0.0016	1.6480
ENSRNOG00000019383	Tef	0.0007	0.7306	ENSRNOG00000017241	RGD1307648	0.0016	1.8322
ENSRNOG00000008787	NP_001101441.1	0.0008	0.8095	ENSRNOG00000012258	Rras2	0.0019	1.2015
ENSRNOG00000013811	Lins2_predicted	0.0009	0.5008	ENSRNOG00000011007	Ube2o_predicted	0.0023	1.3618
ENSRNOG00000020216	Gmpr2	0.0012	0.7462	ENSRNOG00000005083	RGD1311072	0.0023	1.6224
ENSRNOG00000012021	Ctnnbl1	0.0014	0.5647	ENSRNOG00000020289	Akt1s1_predicted	0.0023	1.5188
ENSRNOG00000010352	Dnajc3	0.0014	0.7643	ENSRNOG00000005713	MGC94183	0.0023	1.2412
ENSRNOG00000013596	NP_001099878.1	0.0018	0.7369	ENSRNOG00000001483	NP_001101802.1	0.0023	1.4976
ENSRNOG00000016780	RGD1310951_predicted	0.0022	0.5290	ENSRNOG00000017469	Anxa1	0.0024	1.6118
ENSRNOG00000037707	Armcx6	0.0030	0.1696	ENSRNOG00000020385	Fads3	0.0033	1.7152
ENSRNOG00000015416	RGD1306658	0.0031	0.8099	ENSRNOG00000015294	NP_001100026.1	0.0035	1.4129
ENSRNOG00000024695	NP_001101901.1	0.0033	0.7688	ENSRNOG00000017493	Miz1	0.0035	1.3829
ENSRNOG00000030607	ENSRNOG00000030607	0.0034	0.6941	ENSRNOG00000003983	LOC678975	0.0037	1.3900
ENSRNOG00000013186	RGD1310666_predicted	0.0034	0.8765	ENSRNOG00000029939	Gypc	0.0037	1.5055
ENSRNOG00000018171	St8sia6	0.0035	0.5362	ENSRNOG00000007087	Ebna1bp2	0.0038	1.2800
ENSRNOG00000025937	NP_001102465.1	0.0035	0.6212	ENSRNOG00000014171	Tnfsf13	0.0050	1.7796
ENSRNOG00000008757	RGD1311364	0.0054	0.7079	ENSRNOG00000007664	RGD1560810_predicted	0.0051	2.1199

ENSRNOG00000011382	Wdr33	0.0054	0.8141
ENSRNOG00000018288	Ncoa6	0.0056	0.7380
ENSRNOG00000007744	Ehd3	0.0057	0.8181
ENSRNOG00000020342	RGD1310271_predicted	0.0065	0.4777
ENSRNOG00000003875	NP_001101726.1	0.0065	0.7360
ENSRNOG00000001372	Oas1b	0.0073	0.5183
ENSRNOG00000000640	Egr2	0.0080	0.5060
ENSRNOG00000006822	RGD1311640_predicted	0.0080	0.2974
ENSRNOG00000004474	Klhdc2	0.0090	0.5643
ENSRNOG00000015109	NP_001099556.1	0.0095	0.5758
ENSRNOG00000033556	Spen_predicted	0.0096	0.8510
ENSRNOG00000022303	LOC690214	0.0108	0.6210
ENSRNOG00000031421	Eif1a	0.0109	0.7548
ENSRNOG00000030253	LOC691123	0.0113	0.5261
ENSRNOG00000009347	NP_001102717.1	0.0126	0.8887
ENSRNOG00000023646	Zfp11	0.0136	0.6680
ENSRNOG00000040490	SNORA17	0.0136	0.5906
ENSRNOG00000018714	Arl5b	0.0148	0.5833
ENSRNOG00000021669	NP_001100102.1	0.0167	0.3355
ENSRNOG00000007726	Mcam	0.0167	0.5614
ENSRNOG00000021702	LOC500392	0.0175	0.0988
ENSRNOG00000019308	Arrb2	0.0175	0.4620
ENSRNOG00000004629	NP_001100206.1	0.0175	0.7766
ENSRNOG00000037270	Avpr2	0.0183	0.7019
ENSRNOG00000000237	RGD1310429_predicted	0.0189	0.5000
ENSRNOG00000008155	Dus4l_predicted	0.0203	0.7560
ENSRNOG00000002775	LOC304860	0.0254	0.5577
ENSRNOG00000020936	Nradd	0.0265	0.4057
ENSRNOG00000002671	Nme2	0.0279	0.4789
ENSRNOG00000006583	Ptgds2	0.0308	0.6812
ENSRNOG00000013532	Pgam2	0.0315	0.2738
ENSRNOG00000009481	Ddhd1	0.0321	0.5874
ENSRNOG00000024237	NP_001101858.1	0.0322	0.1533
ENSRNOG00000017477	Mmp23	0.0332	0.2453
ENSRNOG00000001762	Pcyt1a	0.0334	0.7254
ENSRNOG00000012954	Eefsec	0.0369	0.7719
ENSRNOG00000013479	Stard7	0.0407	0.5758
ENSRNOG00000004481	NP_001099445.1	0.0415	0.6085
ENSRNOG00000020464	LOC502522	0.0437	0.3450
ENSRNOG00000004667	NP_001100187.1	0.0439	0.5115
ENSRNOG00000013992	Arfrp1	0.0456	0.8292
ENSRNOG00000025808	NP_001100361.1	0.0476	0.6744
ENSRNOG00000007064	RGD1306209	0.0480	0.7635
ENSRNOG00000026376	4921524L21Rik	0.0495	0.1558
ENSRNOG00000036517	ENSRNOG00000036517	0.0495	0.1638
ENSRNOG00000020000	RGD1560566_predicted	0.0063	1.5374
ENSRNOG00000009340	NP_001102423.1	0.0065	1.3956
ENSRNOG00000019444	Cpsf5	0.0067	1.1519
ENSRNOG00000029490	Zfp768	0.0070	3.2049
ENSRNOG00000030408	LOC294513	0.0070	1.3052
ENSRNOG00000033887	NP_001101511.1	0.0076	1.6222
ENSRNOG00000036683	NP_001100543.1	0.0076	1.2353
ENSRNOG00000018132	LOC361335	0.0080	1.4019
ENSRNOG00000008614	NP_001100213.1	0.0080	1.6778
ENSRNOG00000016346	Prkcd	0.0091	1.2628
ENSRNOG00000030431	MGI:3588187	0.0106	#DIV/0!
ENSRNOG00000022868	Ell3	0.0109	1.9085
ENSRNOG00000003069	Cd38	0.0109	1.1435
ENSRNOG00000001417	Plod3	0.0118	1.6011
ENSRNOG00000018972	Rab18	0.0118	1.2284
ENSRNOG00000005153	LOC690422	0.0121	1.5365
ENSRNOG00000011137	RGD1560834_predicted	0.0126	3.5077
ENSRNOG00000004649	Il1b	0.0128	1.5600
ENSRNOG00000017847	RGD1562823_predicted	0.0128	1.1655
ENSRNOG00000012724	CA123_RAT	0.0131	1.6108
ENSRNOG00000003398	Tomm40b	0.0167	1.3285
ENSRNOG00000007340	RGD1562214_predicted	0.0167	1.3140
ENSRNOG00000010673	Eral1	0.0175	1.5137
ENSRNOG00000022533	RGD1307875_predicted	0.0177	#DIV/0!
ENSRNOG00000013383	MGC72996	0.0180	1.4449
ENSRNOG00000020178	NP_001099546.1	0.0192	15.7116
ENSRNOG00000005932	NP_001101410.1	0.0192	15.1935
ENSRNOG00000020914	Spnb4	0.0193	6.8867
ENSRNOG00000005177	RGD1304982_predicted	0.0210	3.0361
ENSRNOG00000008873	RGD1562258_predicted	0.0230	1.5938
ENSRNOG00000014568	LOC681867	0.0252	1.2777
ENSRNOG00000020179	LOC365090	0.0257	1.4416
ENSRNOG00000006098	LOC689918	0.0259	3.3534
ENSRNOG00000015385	NP_001100164.1	0.0281	1.3383
ENSRNOG00000019045	RGD1306538	0.0286	1.1996
ENSRNOG00000022150	ENSRNOG00000022150	0.0298	2.7412
ENSRNOG00000011154	Gpr116	0.0308	#DIV/0!
ENSRNOG00000026686	Ddx4	0.0308	#DIV/0!
ENSRNOG00000016617	MGC116096	0.0308	6.5597
ENSRNOG00000016961	Rps27	0.0308	1.9182
ENSRNOG00000019802	NP_001101946.1	0.0329	4.3030
ENSRNOG00000016434	Prkd2	0.0354	2.6256
ENSRNOG00000021442	NP_001101930.1	0.0379	1.4112
ENSRNOG00000009974	Coq3	0.0381	1.3312
ENSRNOG00000017002	Adrb1	0.0382	2.3324
ENSRNOG00000024922	NP_001100600.1	0.0422	8.2048
ENSRNOG00000005433	Shq1	0.0422	8.0389
ENSRNOG00000011158	Ppp2r2a	0.0429	1.2191
ENSRNOG00000028711	Dgat1	0.0430	1.7211
ENSRNOG00000008577	NP_001099324.1	0.0456	1.2643
ENSRNOG00000002627	NP_001100669.1	0.0476	1.1958
ENSRNOG00000020325	RGD1308276	0.0492	1.7240
ENSRNOG00000013231	Ptafr	0.0498	2.3703

* S = susceptible congenic control; R = *Mcs5a* resistant congenic; FC = Fold Change of the average tag counts

Table 3: List of Bacterial Artificial Chromosomes used in the FISH analysis and corresponding 4C tag density of the covered region.

ID in Figure 5B	BAC_ID	Coordinates of genomic insert*	4C tag density (tpm/bp)**	FISH interaction frequency***	n****
1	CH230-411H18	chr4:159,361,134-159,519,926	0.02599956	0.0208333	432
2	CH230-171F5	chr7:117,070,861-117,305,522	0.0214181	0.02760736	326
3	CH230-161G22	chr1:43,133,482-43,360,116	0.00430156	0.01269036	788
4	CH230-376A10	chr2:34,225,211-34,419,932	0.00349723	0.0102489	683
5	CH230-254J4	chr15:43,697,259-43,987,044	0.00183534	0.01383399	506
6	CH230-213N13	chr1:245,916,442-246,135,011	0.00175445	0.01622419	678
7	CH230-165G1	chr17:67,954,222-68,048,223	0.00033186	0.00289017	346
fixed	CH230-298P15	chr5:61,589,370-61,739,516	*Mcs5a* BAC	*Mcs5a* BAC	

* BAC insert coordinates rat genome assembly Nov. 2004 (Baylor 3.4/rn4) in UCSC Genome Browser.

** 4C tag density is calculated by the sum of the tag count in the BAC region adjusted for the total amount of basepairs of the mappable portion of the BAC region.

*** The FISH interaction frequency is calculated by counting the number of overlap signals between a variable BAC and the *Mcs5a* BAC (CH230-298P15).

**** n=number of alleles analyzed.

Table 4: Summary of the chi-square tests for independent distributions of the observed and expected hit rate/non-hit rate of groups of genes overlapping with *Mcs5a*-associated fragments.

	EX vs NONEX	EXDE vs EXNONDE	*Mcs5a*-over vs *Mcs5a*-under	EX vs NONEX, 4C DIFF
n 1	11354	198	95	198
n hit 1	2488	45	17	11
n nothit 1	8866	153	78	187
Percentage of features hit 1	21.9%	22.8%	17.9%	5.6%
target size 1	538312573	9043514	4346741	9043514
target size hit 1	202701485	2915728	1405569	1127492
target size non-hit 1	335611088	6127786	2941172	7916022
AVG Featurespan 1	47411.71156	45674.31313	45755.16842	45674.31313
AVG Featurespan hit 1	81471.65796	64793.95556	82680.52941	102499.2727
AVG Featurespan non-hit 1	37853.72073	40050.88889	37707.33333	42331.66845
Hit Featurespan correction 1	1.718386772	1.418608209	1.807020546	2.244133862
Non-hit Featurespan correction 1	0.798404434	0.876879938	0.824110907	0.926815655
n hit corrected 1	1447.869619	31.72123192	9.407751362	4.901668383
n non-hit corrected 1	11104.64775	174.4822675	94.64745505	201.7661214
n hit adjusted OBSERVED 1	1309.626681	30.4592499	8.589059694	4.696089027
n non-hit adjusted OBSERVED 1	10044.37332	167.5407501	86.41094031	193.303911
n adjusted 1	11354	198	95	198
n hit adjusted EXPECTED 1	770.497722	22.85971556	14.92370278	6.415157793
n non-hit adjusted EXPECTED 1	10583.50228	175.1402844	80.07629722	191.5848422
n adjusted expected 1	11354	198	95	198
n 2	15834	11156	103	11156
n hit 2	2009	2443	28	811
n nothit 2	13825	8713	75	10345
Percentage of features hit 2	12.7%	21.9%	27.2%	7.3%
target size 2	613673050	529269059	4541174	529269059
target size hit 2	230943362	199785757	1510159	81738623
target size non-hit 2	382729688	329483302	3031015	447530436
AVG Featurespan 2	38756.66604	47442.54742	44089.06796	47442.54742
AVG Featurespan hit 2	114954.3863	81778.86083	53934.25	100787.4513
AVG Featurespan non-hit 2	27683.8834	37815.13853	40413.53333	43260.55447
Hit Featurespan correction 2	2.966054566	1.723745146	1.223302113	2.124410614
Non-hit Featurespan correction 2	0.714299919	0.797072261	0.916633878	0.911851425
n hit corrected 2	677.3307622	1417.262874	22.88886752	381.7529411
n non-hit corrected 2	19354.61511	10931.25483	81.82110853	11345.04999
n hit adjusted OBSERVED 2	535.3875933	1280.395348	22.51507873	363.1710907
n non-hit adjusted OBSERVED 2	15298.61241	9875.604652	80.48492127	10792.82891
n adjusted 2	15834	11156	103	11156
n hit adjusted EXPECTED 2	1074.516552	1287.994883	16.18043564	555.3150737
n non-hit adjusted EXPECTED 2	14759.48345	9868.005117	86.81956436	10613.61911
n adjusted expected 2	15834	11156	103	11168.93418

Chi-square test 2x2 matrix				
n hit OBSERVED 1	1309.626681	30.4592499	8.589059694	4.696089027
n non-hit OBSERVED 1	10044.37332	167.5407501	86.41094031	193.303911
n hit EXPECTED 1	770.497722	22.85971556	14.92370278	6.415157793
n non-hit EXPECTED 1	10583.50228	175.1402844	80.07629722	191.5848422
n hit OBSERVED 2	535.3875933	1280.395348	22.51507873	363.1710907
n non-hit OBSERVED 2	15298.61241	9875.604652	80.48492127	10792.82891
n hit EXPECTED 2	1074.516552	1287.994883	16.18043564	361.4520219
n non-hit EXPECTED 2	14759.48345	9868.005117	86.81956436	10794.54798
c11 Chi2	377.2367208	2.526405987	2.688857021	0.460658571
c12 Chi2	27.46350181	0.329752361	0.501118363	0.015425006
c21 Chi2	270.5030774	0.044839404	2.480013758	0.008175905
c22 Chi2	19.69310342	0.005852543	0.462196549	0.000273768
Chi2 statistic	694.8964034	2.906850294	6.132185691	0.48453325
P-value	3.8505E-153	0.088203997	0.013274268	0.486376086

KEY RESEARCH ACCOMPLISHMENTS

Training:

- Actively participated in the Gould lab meeting/journal club (SoW Task 1)
- Actively participated in McArdle Lab student/postdoc seminar series (SoW Task 1)
- Attended seminar series on a variety of cancer biology and related topics (SoW Task 1)
- Visited Dr. Job Dekker's lab (UMass Medical School, Worcester, MA) to learn the 3C technology (SoW Task 2)
- Presented at an international scientific meeting: Keystone Symposia 'Complex Traits: Biological and Therapeutic Insights', Santa Fe, NM (SoW Task 3)
- Presented at an international scientific meeting: Keystone Symposia 'Chromatin Dynamics and Higher-Order Organization', Coeur D'Alene, ID (SoW Task 3)
- Presented a poster at the Era of Hope DoD BCRPM meeting, Baltimore, MD (SoW Task 3)
- Presented at an international scientific meeting: 'Rat Genomics and Models', Cold Spring Harbor, NY (SoW Task 3)
- Regular discussions with members of the mentoring committee (SoW Task 4)

Research:

- Completed the 3C experiments (SoW Task 1)
- Concluded 3C experiments, generated a model as a hypothesis for regulation of the *Fbxo10* gene by the *Mcs5a* locus in rats and humans (SoW Task 2)
- Established the luciferase assay (SoW Task 3a)
- Screened (nearly) all breast cancer polymorphisms in *MCS5A1* and *MCS5A2* for promoter activity (SoW Task 3b)
- Identified the promoter element of the *FBXO10* gene containing 1 breast cancer SNP (SoW Task 3b)
- Completed Luciferase assay screen for transcriptional regulatory properties of the *MCS5A2* SNPs on the promoter element of the *FBXO10* gene (SoW Task 3c).
- Completed computational prediction of TFBS on functional SNPs (SoW Task 5a).
- Established ChIP assay showing CTCF binding to the interacting 3C elements (SoW Task 5c).

REPORTABLE OUTCOMES

- Abstract Keystone Symposia ‘Complex Traits: Biological and Therapeutic Insights’, Oral and Poster presentation

- Travel Scholarship Award Keystone Symposia ‘Complex Traits: Biological and Therapeutic Insights’

- Abstract Era of Hope DoD BCRPM Meeting, Poster presentation

- Abstract Keystone Symposia ‘Chromatin Dynamics and Higher-Order Organization’, Poster presentation

- Travel Scholarship Award Keystone Symposia ‘Chromatin Dynamics and Higher-Order Organization’

- Abstract ‘Rat Genomics and Models’, Oral and Poster presentation

- Manuscript (submitted) entitled: Functional Analysis of a Human/Rat Conserved Breast Cancer Susceptibility Locus Identifies a Non-Mammary Cell-Autonomous Repressive Gene Regulatory Mechanism

CONCLUSION

Over the last five years, GWAS have uncovered genetic variants associated with complex traits, such as breast cancer susceptibility. Nevertheless, the molecular mechanisms through which these alleles elicit their action remain highly elusive. Understanding mechanistically how these alleles interact to modulate breast cancer susceptibility could yield great benefits for population-based screening, and pharmaceutical intervention strategies. This study details a number of mechanisms of action of breast cancer risk-associated alleles of the *Mcs5a/MCS5A* locus, taking advantage of the availability of well-characterized rat models, and human cell- and population-based models. These results reflect the complexity that can be anticipated in studying the mechanisms of many other low-penetrance non-coding complex trait alleles.

The non-protein coding rat mammary carcinoma susceptibility locus *Mcs5a* has previously been shown to consist of two interacting genetic elements that must be located on the same chromosome to elicit mammary carcinoma resistance. Our group showed previously that the expression level of genes located within 1 Mb surrounding the locus were not differentially expressed in the mammary gland of susceptible congenic control and *Mcs5a* resistant congenic rats. The two genes directly neighboring the locus, however, were found to be differentially expressed in the immune system, namely *Fbxo10* in the thymus, and *Frmpd1* in the spleen. Here, I show that the transcript level regulation of *Fbxo10* in the thymus by the *Mcs5a* resistant allele, but not that of *Frmpd1* in the spleen, requires the same genetic synthetic interaction that is essential for the mammary carcinoma resistance phenotype. The differential *Fbxo10* transcript level has also been demonstrated in CD4+CD8- thymocytes, CD8+CD4- thymocytes,

CD4+CD8+ thymocytes, and CD3+ splenic T-lymphocytes. This result identifies *Fbxo10* as a strong candidate for driving the mammary carcinoma resistance phenotype likely modifying the resistance phenotype via activities in T-lymphocytes. This hypothesis is further substantiated by the observation that the carcinoma susceptibility phenotype is not transferable from donor to recipient animals in a mammary gland transplantation assay. In contrast, the transplanted mammary glands adopt the host phenotype, suggesting a carcinoma development control mechanism that is not mammary cell-autonomous, but instead act abscopally through cell types/tissues which may modulate mammary carcinogenesis, such as the immune system.

Detailed analysis of the *Fbxo10/FBXO10* transcript level regulatory mechanism mediated by the *Mcs5a/MCS5A* locus revealed striking similarities between rats and humans. First, the TSSs of the *Fbxo10/FBXO10* transcripts were found to be located in the CpG island in *Mcs5a1/MCS5A1* in both rat and human thymus and spleen RNA samples. The human *FBXO10* TSS cluster was found to be located 150 bp away from a breast cancer risk-associated SNP (rs6476643). Second, both the rat and human *Mcs5a/MCS5A* locus display a similar pattern of higher-order chromatin structure that is likely mediated by CTCF/cohesin binding. This structure is thought to form an insulator loop, likely spatially and functionally isolating *Tomm5/TOMM5*. In the human, the *MCS5A1* and *MCS5A2* breast cancer risk-associated polymorphisms are located at both sides of the loop, and thus are brought into closer physical proximity. Third, functional assessment of the transcriptional regulatory properties of the resistant (R) and susceptible (S) allele of rs6476643 in combination with the R and S allele of each of the 15 *MCS5A2* risk-associated SNPs, revealed an overall significantly lower transcriptional activity for the RR combinations compared to the SS combinations, thus reflecting the resistant *Mcs5a1-Mcs5a2* interaction controlling thymic *Fbxo10* transcript levels and mammary carcinoma resistance in the rat. As the R allele of multiple *MCS5A2* SNPs combined with the R allele of rs6476643 can independently lower transcriptional activity compared to the combined SS alleles, it can be anticipated that potential *FBXO10* transcript level regulation in a T-lymphocyte population involves multiple interactions between the *MCS5A1* and *MCS5A2* breast cancer-associated polymorphisms, which are separated by 60 Kb. Physical interactions between the fragments containing the *MCS5A1* and *MCS5A2* risk-associated polymorphisms, however, were not directly observed, likely due to limited sensitivity of the 3C assay and the transient, stochastic and/or structurally diffuse nature of the putative interactions.

Transcriptional gene regulation is currently considered to be a process under control of local and global nuclear mechanisms, such as (allele-specific) RNA polymerase II and transcription factor binding, and colocalization via higher-order chromatin looping of coregulated genes in transcription factories. Without prior knowledge of the transcription factors that give rise to the *Fbxo10* transcript level regulation, we explored in which nuclear environment the *Mcs5a* locus is situated. The 4C assay revealed that the genomic fragments that likely physically associate with the *Mcs5a* locus are non-randomly distributed over the chromosomes and are located near or within genes. Interestingly, the location of the *Mcs5a*-associated fragments was found to relate to the global gene expression profile. Expressed gene regions are enriched in these fragments compared to non-expressed gene regions. Strikingly, genes expressed at a lower level are enriched with these fragments, compared to genes expressed at a higher level in the *Mcs5a* resistant congenic line, suggesting a global repressive mechanism of gene expression regulation that complement the local repressive expression regulation of *Fbxo10*. Similar mechanisms of transcriptional coregulation via long-distance chromatin looping have been described in genome-wide studies. These findings confirm the role of the resistant *Mcs5a* allele as a transcriptional repressor. The functional characterization

of the *Mcs5a* gene regulatory mechanisms implicates that the (down)regulation of the transcript level of *Fbxo10* in the T-lymphocytes is a strong candidate mechanism of action of this non-coding breast cancer susceptibility locus.

By finely mapping QTL in rodent models we can assure that the causative polymorphisms are within the mapped interval, as is the case for *Mcs5a*. The difficulties in identifying causative polymorphisms from GWAS as well as their functions could be due to limited structural and mechanistic knowledge regarding the *cis*-acting polymorphisms and their possible interactions within a locus (e.g. synthetic interactions). Furthermore, the lack of knowledge of which cell type(s) to use for functional studies can compound these issues. Extending high-resolution comparative genomic studies to additional alleles, cell types, and traits will provide useful models to help interpret the results of GWAS findings.

REFERENCES

Bell, A. C., West, A. G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387-396.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M.*, et al.* (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol *18*, 630-634.

Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. Nat Methods *3*, 17-21.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410-413.

Holloway, A. F., and Oakford, P. C. (2007). Targeting epigenetic modifiers in cancer. Curr Med Chem *14*, 2540-2547.

Jongeneel, C. V., Iseli, C., Stevenson, B. J., Riggins, G. J., Lal, A., Mackay, A., Harris, R. A., O'Hare, M. J., Neville, A. M., Simpson, A. J., and Strausberg, R. L. (2003). Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing. Proc Natl Acad Sci U S A *100*, 4702-4705.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Samuelson, D. J., Aperavich, B. A., Haag, J. D., and Gould, M. N. (2005). Fine mapping reveals multiple loci and a possible epistatic interaction within the mammary carcinoma susceptibility quantitative trait locus, Mcs5. Cancer Res *65*, 9637-9642.

Samuelson, D. J., Haag, J. D., Lan, H., Monson, D. M., Shultz, M. A., Kolman, B. D., and Gould, M. N. (2003). Physical evidence of Mcs5, a QTL controlling mammary carcinoma susceptibility, in congenic rats. Carcinogenesis *24*, 1455-1460.

Samuelson, D. J., Hesselson, S. E., Aperavich, B. A., Zan, Y., Haag, J. D., Trentham-Dietz, A., Hampton, J. M., Mau, B., Chen, K. S., Baynes, C.*, et al.* (2007). Rat Mcs5a is a compound quantitative trait locus with orthologous human loci that associate with breast cancer risk. Proc Natl Acad Sci U S A *104*, 6299-6304.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive

chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet *38*, 1348-1354.

Smits, B. M., Samuelson, D. J., Guryev, V., Chung, L. M., Wiederholt, C. J., Bruijn, E., Traun, B., DeVries, T. L., Mau, B., Haag, J. D., Cuppen, E., Newton, M. A., and Gould, M. N. (2010) submitted

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev *20*, 2349-2354.

t Hoen, P. A., Ariyurek, Y., Thygesen, H. H., Vreugdenhil, E., Vossen, R. H., de Menezes, R. X., Boer, J. M., van Ommen, G. J., and den Dunnen, J. T. (2008). Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res *36*, e141.

APPENDICES

none